

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 28, 2004, 10:34:22 ; Search time 27.5 Seconds

(without alignments)  
9883.862 Million cell updates/sec

Title: US-09-787-016a-1

Perfect score: 4689

Sequence: 1 ctccgtgcccgtccgcccac.....actcttaaatgatactatctcg 2610

Scoring table:

BLOSOM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame-n2p.model -DRV=xlp  
-Q=/cgn2\_1/USFTO\_spool\_p/US09787016/runat\_28042004\_094328\_26905/app\_query.fasta\_1.2759  
-DB=SwissProt\_42 -QWTF=fastan -SUFFIX=rep -MINMATCH=0.1 -LOOPCT=0 -LOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosom62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09787016 @CGN 1.1.28 @runat\_28042004\_094328\_26905 -NCPU=6 -ICPU=3  
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG  
-DBG TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DLEXT=7

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2770	59.1	1225	DAT1 HUMAN	Q9btc0 homo sapien
2	2156.5	46.0	614	DAT1 MOUSE	Q8c959 mus musculus
3	232	4.9	2039	PHF2 HUMAN	Q92576 homo sapien
4	203.5	4.3	656	CGBP HUMAN	Q9p0u4 homo sapien
5	194.5	4.1	660	CGBP MOUSE	Q9cww7 mus musculus
6	193	4.1	483	CYL2 BOVIN	Q28092 bos taurus
7	189.5	4.0	705	TRDN RABIT	Q28920 cryptolagus
8	188.5	4.0	1010	HC58 MOUSE	Q8b5g0 mus musculus
9	182.5	3.9	704	NP14 RAT	P41777 rattus norv
10	182	3.9	699	NP14 HUMAN	Q14978 homo sapien
11	175.5	3.7	2004	MY53 HUMAN	Q92794 homo sapien
12	174.5	3.7	1012	HC58 HUMAN	Q9bw10 homo sapien
13	174	3.7	700	TRDN CANFA	P82179 canis famil
14	173	3.7	755	PRR1 HUMAN	Q92766 homo sapien
15	172	3.7	5038	PCLOA MOUSE	Q9qx77 mus musculus
16	170	3.6	1692	RIM1 HUMAN	Q84ur5 homo sapien
17	169	3.6	728	TRDN HUMAN	Q13061 homo sapien
18	167.5	3.6	1087	NFH MOUSE	P19246 mus musculus

19	167.5	3.6	2476	1	ATRX MOUSE	Q61687 mus musculus
20	167.5	3.6	3644	1	MINT MOUSE	Q62504 mus musculus
21	167	3.6	1101	1	PHF2 HUMAN	Q75151 homo sapien
22	165.5	3.5	1359	1	ATRX CAEBL	Q9u7e0 caenorhabdi
23	163.5	3.5	1394	1	CNG4 BOVIN	Q28181 bos taurus
24	163.5	3.5	1781	1	AK12 HUMAN	Q02952 homo sapien
25	163	3.5	1096	1	PHF2 MOUSE	Q9wtu0 mus musculus
26	162.5	3.5	1805	1	NEST RAT	P21263 rattus norv
27	161	3.4	2492	1	ATRX HUMAN	P46100 homo sapien
28	161	3.4	2492	1	ATRX PANTR	Q7yqm4 pan troglod
29	161	3.4	2492	1	ATRX PONPY	Q7yqm3 pongo pygma
30	159	3.4	5085	1	PCLO RAT	Q91ks6 rattus norv
31	158.5	3.4	679	1	RRP1 DROME	P27864 drosophila
32	157	3.3	713	1	ICAL PIG	P12675 sus scrofa
33	156.5	3.3	778	1	TAST HUMAN	Q12815 homo sapien
34	156	3.3	881	1	PHPE MOUSE	Q94dh9 mus musculus
35	156	3.3	1861	1	MAP2 RAT	P15146 rattus norv
36	155.5	3.3	1234	1	JMJ MOUSE	Q62315 mus musculus
37	155	3.3	5147	1	PCLO HUMAN	Q9y6v0 homo sapien
38	154.5	3.3	1618	1	NEST HUMAN	P48681 homo sapien
39	154	3.3	852	1	SRCH RABIT	P16230 cryptolagus
40	154	3.3	1026	1	NPH HUMAN	P12036 homo sapien
41	154	3.3	1191	1	CING MOUSE	P59242 mus musculus
42	153.5	3.3	1338	1	ACIN MOUSE	Q9jix8 mus musculus
43	153.5	3.3	1560	1	SMCX HUMAN	P41229 homo sapien
44	153.5	3.3	6632	1	UN89 CAEBL	O01761 caenorhabdi
45	153	3.3	952	1	YK15 CAEBL	P46012 caenorhabdi

## ALIGNMENTS

RESULT 1  
DAT1 HUMAN  
ID DAT1 HUMAN STANDARD; PRT: 1225 AA.  
AC Q9BTC0: O15043; Q9BOW0; Q9BW03; Q9H4G6; Q9NTU8; Q9NUM8;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DE Death associated transcription factor 1.  
GN DAT1 OR KIAA0333.  
OS Homo sapiens (Human)  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
CX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 3).  
RC TISSUE=Placenta;  
RA Isoqai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,  
RA Watanabe M., Hosoi T., Raku Y., Kodaira H., Kondo H., Sugawara M.,  
RA Takahashi N., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,  
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,  
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,  
RA Ninomiya K., Iwayanagi T.,  
RT "NEBO human cDNA sequencing project."  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
MEDLINE=2163749; PubMed=11780052;  
RA Delucas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,  
RA Jones M., Stavriles G., Almeida J.P., Babbage A.K., Bagguley C.L.,  
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,  
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,  
RA Buck D., Burfill W.D., Butler A.P., Carder C., Carter N.P.,  
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,  
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,  
RA Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,  
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,  
RA Graham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,  
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,  
RA Huckle B., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
RA Kay M.P., Kimberley A.M., King A., Knights K., Laird G.K., Lawlor S.,  
RA Lehaeslaio M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,



Db 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120  
QY 625 ACCAGAGCGGCCCGCCAGTCTGCTCCACAGCTGTGAAGAGACGACGAGCTCTCTCAA 684  
Db 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140  
QY 685 AAGGTGAAGAGGGGATGACACACGATGACACCTCCGATAGTGAAGAGCGATGCGCTGACC 744  
Db 141 LysValLysGlyGlyAspAspHisAspAspThrSerAspSerAspSerAspGlyLeuThr 160  
QY 745 TTGAAGAGCTTCAGAAATCGCTTCGACAGGAGCGGGAACAGGAGCCCACTGAGAGGCC 804  
Db 161 LeuLysGluLeuGlnAsnArgLeuArgArgLysGluGlnGluProThrGluArgPro 180  
QY 805 CTGAAGAGGATCCAGAGTCCGCTCGGAGAGAGCGCCGGAGGAGGGTCCCGCCGAGACT 864  
Db 181 LeuLysGlyIleGlnSerArgLeuArgLysLysArgArgGluGluGlyProAlaGluThr 200  
QY 865 GTGGCTCGAGGCCAGTGAACACTGTGAGGCGGCTCCGACGAGTAAGCAGGAGCCGAG 924  
Db 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGlu 220  
QY 925 AACGATCAGGCGGTTGTCTCCAGGCTGGGAAAGATGACAGAGAGTAAGTTGGAGGGA 984  
Db 221 AsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGluGly 240  
QY 985 AAGCGGCTCAGGACATCAAGATCAGAGCGCTGAGACTGGGCGGACCGACGCTCAA 1044  
Db 241 LysAlaAlaGlnAspIleLysAspGluGluProGlyAspLeuGlyArgProLysProGlu 260  
QY 1045 TGTGAGGTTACAGCCCAAGCCCTGATTGTCATTTGCCGCGACCTCCACAAACAGG 1104  
Db 261 CysGluGlyTyArgProAsnAlaLeuTyCysIleCysArgGlnProHisAsnArg 280  
QY 1105 TTATGATTGCTGAGCGCTGTGAAGATGGTTTCATGCGGATGTGTGGCGATTCT 1164  
Db 281 PheMetIleCysCysAspArgCysGluGluThrPheHisGlyAspCysValGlyLys 300  
QY 1165 GAGGCTCAGAGGAGCTTTTGGAGAGGATGGGAGAGCTATATCTGCCCAACTGCACC 1224  
Db 301 GluAlaAspGlyArgLeuGluArgGlnGlyAspTyIleCysProAsnCysThr 320  
QY 1225 ATTCTGCAAGTCCAGGATGAGACTCATTTCAGAAACGGCAGATCAGCAGAGCTAAATGG 1284  
Db 321 IleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLysTrp 340  
QY 1285 AGACTGAGAGTCTGATGGCCAGCTGATCTACAGTATAGGACATAGACAGAGTCT 1344  
Db 341 ArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSer 360  
QY 1345 AGCGAAGACCAAGGATAAAGGTAGAATTGAGAAAGCTGCAAATCCAAAGTGGCAAGAG 1404  
Db 361 SerGluAspGlnGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLysLys 380  
QY 1405 AAATCAAGTCTCCAGCTT----- 1425  
Db 381 LysLeuLysIlePheGlnProGlyProGlyProValProThrGlnLeuProValLeuTrp 400  
QY 1425 ----- 1425  
Db 401 GlnValLeuGluIleAlaValSerArgSerIleSerAlaPheThrLeuLeuHisCysIle 420  
QY 1426 -----GTGATAGAGCGCGCTGTGCTCCAAATGATATGCGCCCGGTGCTGTAC 1476  
Db 421 SerCysLysValIleGluAlaProGlyAlaSerLysCysIleGlyProGlyCysCysHis 440  
QY 1477 GTGGCGGACCGGCTCGGTGTACTGAGTAACTGCTATCTCCTCAACACCGCCGAGCG 1536  
Db 441 ValAlaGlnProAspSerValTyrcysSerAsnAspCysIleLeuLysHisAlaAla 460  
QY 1537 ACAATGAAGTTCTTAAGTTCAGGTAAAGACAGAGCCCAAGCCCTAAAGAAAGATGAAG 1596  
Db 461 ThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysProLysGluLysMetLys 480

QY 1597 ATGAAGCCAGAGAGCCCACTCTTCCGAAATGCGGTCTCTCAGGAGGTATTAAATCTCT 1656  
Db 481 MetLysProGluLysProSerLeuProLysCysGlyAlaGlnAlaGlyIleLysIleSer 500  
QY 1657 TCTGTGCACAGACACGACCTCCAGAAAAAAGAGACGACGAGTGAAGAGCGAGTGTG 1716  
Db 501 SerValHisLysArgProAlaProGluLysGluThrValLysLysAlaValVal 520  
QY 1717 GTCCCTCGGCGAGTGAAGCACTCGGGAAGCAAGCAGCTTGTGAGAGCAGCAGCCGTCG 1776  
Db 521 ValProAlaArgSerGluAlaLeuGlyLysGluAlaCysGluSerSerThrProSer 540  
QY 1777 TGGCGAGCGATCACAATTACATGCAAGTAAAGCCAGAAAGACTGCTCTCCCTCCGCG 1836  
Db 541 TrpAlaSerAspHisAsnTyAsnAlaValLysProGluLysThrAlaAlaProSerPro 560  
QY 1837 TCACTGTTGTATAAA 1851  
Db 561 SerLeuLeuTyrls 565

RESULT 2  
DAT1 MOUSE  
ID DAT1 MOUSE STANDARD; PRT; 614 AA.  
AC Q8CB9; Q8BRG2; Q9WV00;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Death associated transcription factor 1 (Death inducer-obliterato-1)  
DE (DIO-1).  
GN DATF1 OR DIO1 OR DIO1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OK NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
RC TISSUE=Pre-B cell;  
RX MEDLINE=99324176; PubMed=10393935;  
RA Garcia-Domingo D., Leonardo E., Grandien A., Martinez P., Albar J.P.,  
RA Izpisua-Belmonte J.-C., Martinez-A C.;  
RT "DIO-1 is a novel gene involved in onset of apoptosis in vitro, whose  
RT misexpression disrupts limb development."  
RL Proc. Natl. Acad. Sci. U.S.A. 96:7992-7997(1999).  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Embryo, and Thymus;  
RX MEDLINE=22354683; PubMed=1246851;  
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
RA Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
RA Baldairelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
RA Gaasterland T., Gariboldi M., Gesi C., Godzik A., Gough J.,  
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
RA Petrovsky N., Pallai R., Pontius J.U., Qi D., Ramachandran S.,  
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
RA Sandelin A., Schneider C., Sempie C.A., Setou M., Shimada K.,  
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
RA Wilming L.G., Wyshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
RA Yuan Z., Zavalan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
RA Hirozane-Kishikawa T., Kono H., Nakamura M., Sakazume N., Sato K.,  
RA Shiraki T., Waki K., Kawai J., Alzawa K., Arakawa T., Fukuda S.,  
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,

RA Birney B., Hayashizaki Y.;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RL 60,770 full-length cDNAs";  
 CC Nature 420:563-573(2002).  
 CC -!- FUNCTION: Putative transcription factor, weakly pro-apoptotic when  
 CC overexpressed.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic; translocates to the nucleus  
 CC after pro-apoptotic stimuli.  
 CC -!- TISSUE SPECIFICITY: Ubiquitous. Expressed at intermediate levels.  
 CC -!- INDUCTION: Upregulated during apoptosis.  
 CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.  
 CC -----  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AJ238332; CAB48401.1; -;  
 DR EMBL: AK042474; BAC31270.1; -;  
 DR EMBL: AK044919; BAC32141.1; -;  
 DR MGD: MGI:1344352; Dtd01.  
 DR GO: GO:0005634; C:nucleus; IDA.  
 DR GO: GO:0006915; P:apoptosis; IDA.  
 DR InterPro: IPR001965; Znf\_PHD.  
 DR Pfam: PF00628; PHD; 1.  
 DR SMART: SM00249; PHD; 1.  
 DR PROSITE: PS01359; ZF\_PHD\_1; 1.  
 DR PROSITE: PS50016; ZF\_PHD\_2; 1.  
 DR Apoptosis; Nuclear protein; Zinc-finger.  
 KW DOMAIN 162 170 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 182 190 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT ZN FING 265 319 PHD-TYPE.  
 FT CONFLICT 45 45 V -> A (IN REF. 1).  
 FT CONFLICT 331 331 D -> N (IN REF. 1).  
 FT CONFLICT 353 353 I -> V (IN REF. 1).  
 SQ SEQUENCE 614 AA; 67409 MW; 15ACC5A4BAEF3539 CRC64;

## Alignment Scores:

Pred. No.: 1,28e-113 Length: 614  
 Score: 2156.50 Matches: 450  
 Percent Similarity: 76.80% Conservative: 40  
 Best Local Similarity: 70.53% Mismatches: 113  
 Query Match: 45.99% Indels: 36  
 DB: 1 Gaps: 12

US-09-787-016a-1 (1-2610) x DAT1\_MOUSE (1-614)

QY 265 ATGACGACAAAGCGACCGACCAATGAGAGCGCACCTAAGCCCATCAAGCCACCCAGC 324  
 DB 1 MetaspaspysglyhisleuSeranGluGluAlaProlysAlaileysProThrSer 20  
 QY 325 AARGAGTTTCAGGAAACATCGGGTTTTCGAGGACCACTATCCAGCGAGCGAGCGCGCA 384  
 DB 21 LysGluPheArglysthrTrpGlyPheArgThrThrileAlaLysargGluGlyAla 40  
 QY 385 GCGAGCGCGAGGCTGACCCACTGGAGCGCCACCCACCCAGCAGCGCTGGCGCTGCC 444  
 DB 41 GlyAspThrGluValaspProSerGluGlnGlnPro-----GlnGlnHisasnLeuSer 58  
 QY 445 CTGCGGCGAGTGGAGGCGACCCAAAGCCACTGAGCGGTGGAGCAGTTCCTGACCAT 504  
 DB 59 LeuArgSerGlyArgGlnProLysArgThrGluArgValGluGluPheLeuThrThr 78  
 QY 505 GCGCGGCGCGCGAGGAGCAGTCCTGCTCCCTGAGGATTCCTGGTGGAGCCGCG 564  
 DB 79 ValArgArgGlyLysLysAsnValProValSerLeuGluaspSerGluProThr 98  
 QY 565 TCCTGCGCCCGCACAGCGCGAGACAGCCCTCCGAGCGGACCGCTGGAAAGCGCTTCTGAG 624  
 DB 99 SerSerThrValThrAspValGluThrAlaSerGluGluGlySerValGluSerSerSerGlu 118

QY 625 ACCAGAGCGCGCCCGCAGCTCTTCACAGCTGTGAAGGAACGACCCAGCCCTCTTCTGAA 684  
 DB 119 IleArgSerGlyProValSerAspSerleugly--LysGluHisProAlaSerSerGlu 137  
 QY 685 AAGGTGAAAGAGGAGGATGACCATGATGACCTCCGATAGTACGACGAGGAGCGCTGAC 744  
 DB 138 LysAlaLysGlyGluGluGluAspThrSerAspSerAspSerAspGlyleuThr 157  
 QY 745 TTGAAAGAGCTTCAGAAATCGCCTTCGACAGGAGCGGGAACAGGAGCCCACTGAGAGGCC 804  
 DB 158 LeuLysGluLeuGlnAsnArgLeuArgArgLysArgGluGlnGluProValGluArgSer 177  
 QY 805 CTGAAAGGATTCAGAGTCGCTCGGGAAGAGCCCGGAGGAGGCTCCCGCCGAGCT 864  
 DB 178 LeuArgGlySerGlnAsnArgLeuArgLysLysArgGluGluAspSerAlaGluThr 197  
 QY 865 GTGGCTCCGAGGCGCAGTGACACTGTGAGGCGCTCTCTGCC--AGTAAGCAGGAGGCC 921  
 DB 198 --GlySerValGlnIleGlySerAlaGlnAspArgProLeuCysLysGlnGluPro 216  
 QY 922 GAGAACGATCAGGGGGTTGTCTCCAGCTCGGGAAGAGATGACAGAGAGAGTAAGTTGAG 981  
 DB 217 GluAlaSerGlnGlyProValSerGlnSerGlnSerGlnThrAspAspIleGluAsnGlnLeuGlu 236  
 QY 982 GGAAGGCGGCTCAGGACATCAAGATGAGGAGCGCTGGAGACTTGGCGCGCCGAGGCT 1041  
 DB 237 GlyLysAlaThrGlnGlyAsnThrGluGluAsnProArgGluAlaGlyLysProLysPro 256  
 QY 1042 GAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCAATTCGCCCGCAGCTCACAAAC 1101  
 DB 257 GluCysGluValTyAspProAsnAlaLeuTyCysIleCysArgGlnProHisasn 276  
 QY 1102 AGGTTTANGATTCGTGACCGCTGAGAGAGTGTTCATGGGATTCGTGGGCTT 1161  
 DB 277 ArgPheMetIleCysCysAspArgCysGluGlnThrPheHisGlyAspCysValGlyIle 296  
 QY 1162 TCTGAGGCTCGAGGAGGCTTTTGAAGAGGAATGGGAGAGACTATATCTGCCCAACTGC 1221  
 DB 297 SerGluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyIleCysProasnCys 316  
 QY 1222 ACCATTCTGAGTCAGCATGAGATGAGATCTATTCAAGAACCGGACATCAGCAGAACTAA 1281  
 DB 317 ThrIleLeuGlnValGlnAspGluThrAsnGlySerAlaThrAspGluGlnAspSerGly 336  
 QY 1282 TGGAGACCTGGAGATGCTGTGGCCACCATGTGTCAAGTATAGGAACAATAGACAGAG 1341  
 DB 337 CysArgSerValGlyAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLys 356  
 QY 1342 TCTAGCGAAGACCAAGGGATAAAGGGTAGAATTGAGAAAGCTGCAATCCAGTGGCAG 1401  
 DB 357 SerGlyGluAspGlnGlyIleLysGlyArgIleGluLysAlaAlaasnProSerGlyLys 376  
 QY 1402 AAGAACTCAAGATCTCCAGCTGTGATAGAGGCGCTGGTGCTCAAAATGTATTGCT 1461  
 DB 377 LysLysLeuLysIlePheGlnProValValGluAlaProGlyAlaProLysCysIleGly 396  
 QY 1462 CCGCGGTGTCTCAGCTGGCGGAGCCGACTCGGTGTACTGCGAGTAATGACTGTATCTCTC 1521  
 DB 397 ProGlyCysSerSerValAlaGlnProAspSerValTyCysSerAsnAspCysIleLeu 416  
 QY 1522 AAACGCGCGCAGCCACATGAGTTTCTAGCTCAGTGAAGAAAGACAGAGCAAGCCCT 1581  
 DB 417 LysHisAlaAlaAlaThrMetArgPheLeuSerSerGlyLysGluGlnLysThrLysPro 436  
 QY 1582 AAAGAAAGATGAAGATGAAGCCAGAGAGCCAGTCTTCCGAAATGGCGTCTCAGGCA 1641  
 DB 437 LysGluLysValLysThrLysProGluLysPheSerLeuProLysCysSerValGlnVal 456  
 QY 1642 GGTATTAATCTCTCTGTCGACAGAGACAGCTCCAGAGAAAAAGAGACACACAGTG 1701  
 DB 457 GlyIleLysIleSerSerValHisLysArgLeuAlaSerGluLysArgGluAsnProVal 476



CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AF091622; RAF21292.1; -.  
 CC EMBL; D87685; BAAL3438.2; ALT\_INIT.  
 CC EMBL; AL050329; CAB99358.1; -.  
 CC Genew; HGNC:8921; PHF3.  
 CC MIM; 607789; -.  
 CC InterPro; IPR003618; TFS2\_centre.  
 CC InterPro; IPR001965; Znf\_PHD.  
 CC Pfam; PF00628; PHD; 1.  
 CC SMART; SM00249; PHD; 1.  
 CC SMART; SM00510; TFS2M; 1.  
 CC PROSITE; PS01359; ZF\_PHD\_1; 1.  
 CC PROSITE; PS00016; ZF\_PHD\_2; 1.  
 CC Zinc-finger. 717 772 PHD-TYPE.  
 FT ZN FING 929 1029 TFS2M.  
 FT DOMAIN 1797 1865 PRO-RICH.  
 FT SEQUENCE 2039 AA; 229480 MW; FDAFF00576005E9B CRC64;  
 SQ

Alignment Scores:  
 Pred. No.: 9, 15e-06 Length: 2039  
 Score: 232.00 Matches: 94  
 Percent Similarity: 36.43% Conservative: 63  
 Best Local Similarity: 21.81% Mismatches: 154  
 Query Match: 4.95% Indels: 120  
 DB: 1 Gaps: 17

US-09-787-016a-1 (1-2610) x PHF3\_HUMAN (1-2039)

QY 538 TCCTCTGGAGGAT---TCCTGTGAGCCACCTCTCTCCCGCCACAGACGCCGAGACGCC 594  
 Db 595 SerLeuSerAspLysSerHisAlaHisProGlyCysLeuLysGluProHisHisProAla 614  
 QY 595 TCCGAGGCGCAGGTGGAAGAGCCCTTCTGAGACAGCA---AGGGGCCCCAGCTCTGCTCC 651  
 Db 615 GlnThrGlyHisValSerHisSerSerGlnLysGlnCysHisLysProGlnGlnAla 634  
 QY 652 ACAGCTGTGAGGAGACGACCCCTCTCTCTGAAAGGTGAA---GGAGGGGATGACCCAC 708  
 Db 635 ProAlaMetLysThrAsnSerHisValLysGluGluLeuHisProGlyValGluHis 654  
 QY 709 GATGACACCTCCGATAGTGACAGCGATGCGCTTGAAGAGAG---CTT 756  
 Db 655 -----PhelLysGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 670  
 QY 757 CAGAAATCCCTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 816  
 Db 671 GlnProArgGlnArgSerSerLysSerLysSerLysSerLysSerLysSerLysSerLys 690  
 QY 817 CAGACTCCCTCGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 876  
 Db 691 ProAspAsnIleAlaThrIleArgArgGlu-----GlySer----- 702  
 QY 877 GCCAGTGCAGCTGTGGAGGGGCTCTCTGCCAGTAAAGAGGAGGAGGAGGAGGAGGAGGAG 936  
 Db 702 ----- 702  
 QY 937 GTTGTGTCCAGGCTGGGAGAGATGACAGAGAGATGAGTTCGAGGAGGAGGAGGAGGAGGAG 996  
 Db 703 -----AspHisSerSerSerPheGluSerLysTyr----- 712  
 QY 997 GACATCAAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1056  
 Db 713 -----MetTrpThrProSerLysGlnCys---GlyPhe 722  
 QY 1057 GACCCCAAGCCCTGTATGCTATTTGCGGCGAGCCTCTCAACAACAGAGGTTTATGATTC 1116  
 Db 723 -----CysLysLysProHisGlyAsnArgPheMetValGly 734

## RESULT 3

ID PHF3\_HUMAN STANDARD; PRT; 2039 AA.  
 AC Q92576; Q9NQ16; Q9UI45;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE PHD finger protein 3.  
 GN PHF3 OR KIAA0244.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. AND TISSUE SPECIFICITY.  
 RX MEDLINE=21845875; PubMed=11856869;  
 RA Fischer U., Struss A.-K., Hemmer D., Michel A., Henn W.,  
 RA Steudel W.-I., Meese E.;  
 RT "PHF3 expression is frequently reduced in glioma."  
 RL Cytogenet. Cell Genet. 94:131-136(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97191544; PubMed=9039502;  
 RA Nagase T., Seki N., Ishikawa K.-I., Ohira M., Kawarabayashi Y.,  
 RA Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;  
 RT "Prediction of the coding sequences of unidentified human genes. VI.  
 RT The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by  
 RT analysis of cDNA clones from cell line KG-1 and brain."  
 RL DNA Res. 3:321-329(1996).  
 RN [3]  
 RP SEQUENCE OF 83-2039 FROM N.A.  
 RA Matthews L.;  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 CC -!- TISSUE SPECIFICITY: Ubiquitous. Expression is significantly  
 CC reduced or lost in glioblastomas, glioblastoma cell lines,  
 CC anaplastic astrocytomas, and astrocytomas.  
 CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

QY 1117 TGTGACCGCTGTAAGAATGGTTTCATGCGCATTTGTGGCGATTCTTGAGCTCGAGGG 1176
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 735 CysGlyAlaScyAspAspTTPheHisGlyAspCysValGlyLeuSerLeuSerGlnAla 754
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1177 AGGCTTTTGGAAAGAAATGGGGAAGACTATATCTGCCCAAACTGCAACATTCGCAAGT 1236
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 755 GlnGlnMetGlyGluGluAspLysGluTyrValCysValLysCysCys-----Ala 771
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1237 CAGGATGAGACTCATTCAGAAAGCGCAGAT-----CAGCAGGAAGCT 1278
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 772 GluGluAspLysLysThrGluLeuLeuAspProAspThrLeuGluAsnGlnAlaThrVal 791
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1279 AAATGAGACCTCGAGATCTGTCATGCCACCGATTCTCAAGATPATGATAGCAACATGAGCAG 1338
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 792 GluPheHisSerGlyAspLysThr---MetGluCysGluLysLeuGlyLeuSerLysHis 810
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1339 AAGTCTAGCGAA-----GACCAAGGATAAAGGTTAGATTGAGAAAGCT 1383
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 811 ThrThrAsnAspArgThrLysTyrLeuAspAspThrValLysHisLysValLysLeu 830
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1384 GCAAAATCCAAGTGGCAAGAGAAACTCAAGATCTTCCAGCCTGTGATAGAGCGCCTGGT 1443
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 831 LysArgGluSerGlyGluGlyArg-----838
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1444 GCCTCAAAATGTATGGCCCGGCTGTCTACGTGGCGCAGCCGACTCGGTACTGTC 1503
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 839 -----AsnSerSerAspCys 843
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1504 AGTAATGACTGTATCTCAACACGCCCGCAGCGCAATGAACTTTCTAAGCTCAGGTAAA 1563
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 844 ArgAspAsnGluLeuLysLysTrpGlnLeuAlaProLeuArgLysMetGlyGlnProVal 863
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1564 GAACAGAGCCAAAGCCTAAAGAAAGATGAGATGAGCAGCAGAACGCCAGTCTCG 1623
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 864 LeuProArgArgSerSerGluGlyLysSerGluLysLysProLysGluSerThrVal 883
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1624 AAATCGGGTCTCAGCAGGTATTAATCTCTCTGTGTCACAGAGCAGCAGTCCAGAA 1683
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 884 ThrCys-----ThrGlyGluLysAlaSerLysProGlyThrHisGluLysGlnGlu 900
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1684 AAAAAAGACACACAGTGAAGAGGAGCGTGGTGGTCCCTCGCGGAGTGAACATCTCGG 1743
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 901 MetLysLysLysLysValGluLysGlyVal-----LeuAsn 912
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1744 AGGAGAGCGCTGTGAGAGCAGCAGCGCGCTCG 1776
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 913 ValHisProAlaAlaSerAlaSerLysProSer 923
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 4
CGBP_HUMAN STANDARD; PRT; 656 AA.
AC Q9P0U4; Q958C8; Q9P2V7;
DT 28-FEB-2003 (Rel. 41, Created)
DE 10-OCT-2003 (Rel. 41, Last sequence update)
DE Cpg binding protein (Protein containing PHD finger and CXXC domain 1).
GN CGBP OR PCCK1.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=20153771; PubMed=10688657;
RA Shin Voo K., Carlone D.L., Jacobsen B.M., Flodin A., Skalknik D.G.;
RT "Cloning of a mammalian transcriptional activator that binds
RT unmethylated CpG motifs and shares a CXXC domain with DNA
RT methyltransferase, human trithorax, and methyl-CpG binding domain
RL Mol. Cell. Biol. 20:2108-2121(2000).
RN [2]
RP SEQUENCE FROM N.A., AND DNA-BINDING.

```

```

RX MEDLINE=20261256; PubMed=10799292;
RA Fujino T., Hasegawa M., Shibata S., Kishimoto T., Inai S.-I.,
RA Takano T.;
RT "PCCK1, a novel DNA-binding protein with PHD finger and CXXC domain,
RT is regulated by proteolysis.";
RL Biochem. Biophys. Res. Commun. 271:305-310(2000).
RN [3]
RP SEQUENCE FROM N.A.
RX TISSUE=Testis;
RC MEDLINE=21154917; PubMed=11230166;
RA Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,
RA Ansoerg W., Boecker M., Bloeker H., Bauersachs S., Blum H.,
RA Lauber J., Duesterhoeft A., Beyer A., Koehler K., Strack N.,
RA Mewes H.-W., Oettersweider B., Obermaier B., Tampe J., Heubner D.,
RA Wambutt R., Korn B., Klein M., Poustka A.;
RT "Towards a catalog of human genes and proteins: sequencing and
RT analysis of 500 novel complete protein coding human cDNAs.";
RN [4]
RP SEQUENCE FROM N.A.
RX TISSUE=Colon, and Skin;
RC MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold B.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Raha S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Pahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whitling M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzyzanski M.I., Skalska U., Smalus D.B.,
RA Schnerch A., Schein J.B., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RN [5]
RP Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903(2002).
RX DNA-BINDING DOMAIN.
RC MEDLINE=21576202; PubMed=11572867;
RA Lee J.H., Voo K.S., Skalknik D.G.;
RT "Identification and characterization of the DNA binding domain of
RT Cpg-binding protein.";
RL J. Biol. Chem. 276:44669-44676(2001).
CC -!- FUNCTION: Transcriptional activator that exhibits a unique DNA
CC binding specificity for [AC]CpG[AC] unmethylated CpG motifs.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: Ubiquitous.
CC -!- DOMAIN: The acidic domain carries the potential to activate
CC transcription.
CC -!- PTM: May be regulated by proteolysis.
CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC -!- SIMILARITY: Contains 1 CXXC-type zinc finger.
CC -!- SIMILARITY: Contains 1 CXXC-type zinc finger.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF149758; AAF37799.1; -
CC EMBL: AB031069; BAA96307.1; -
CC EMBL: AL136862; CAB66796.1; -
CC EMBL: BC014940; AAL14940.1; -
CC EMBL: BC015733; AAL15733.1; -
CC GO: 0005634; C:nucleus; IDA.

```





CC -!- SUBCELLULAR LOCATION: Calyx; sperm head cytoskeletal structure.  
CC -!- TISSUE SPECIFICITY: Testis.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: Z66789; CAAB6753.1; -;  
CC PIR: I46014; I46014.  
KW Cytoskeleton; Structural protein; Repeat; Sperm; Spermatogenesis.  
FT DOMAIN 25 487 49 X 3 AA REPEATS OF K-K-X.  
FT DOMAIN 189 407 8 X APPROXIMATE TANDEM REPEATS.  
FT REPEAT 189 212 1.  
FT REPEAT 213 239 2.  
FT REPEAT 240 267 3.  
FT REPEAT 268 295 4.  
FT REPEAT 296 323 5.  
FT REPEAT 324 351 6.  
FT REPEAT 352 379 7.  
FT REPEAT 380 407 8.  
SQ SEQUENCE 488 AA; 53562 MW; 5F3107CE2D019C5A CRC64;  
  
Alignment Scores:  
Pred. No.: 0.00114 Length: 488  
Score: 193.00 Matches: 117  
Percent Similarity: 32.99% Conservative: 73  
Best Local Similarity: 20.31% Mismatches: 204  
Query Match: 4.12% Indels: 182  
DB: 1 Gaps: 19  
  
US-09-787-016A-1 (1-2610) x CIVL2\_BOVIN (1-488)  
QY 316 CCACACGACGAGGTTACGAGAAACATGGGGTTTTCGAGGACCACTATCGCCAGCGA 375  
DB 19 ProValSerGluLeuSerLysLysSerTrpAsn----- 29  
QY 376 GAGGGCGCAGGGACCGGAGGCTGACCCACTGGAGCCGCCACCCCGCAGCAGCAGCTG 435  
DB 30 -----GlnGlnHisphe 33  
QY 436 GGCCTCTCCCTGGCGGCGAGTGGAGGAGCCCAAGCGCACTGAGCGCTGGAGCAGTTTC 495  
DB 34 AlaLeuValPheProLysProArgProGlyLysArg----- 46  
QY 496 CTGACCATTTGGCGGCGCGCGCAGGAGAGCATGCTCTCTCTCTCGAGGATTCGTGT 555  
DB 47 -----ArgArgSerLysProSerLeuLeuGlnGlnAsn--- 57  
QY 556 GAGCCACGCTCTGCGCCCGCCGACAGAGCGCGAG----- 588  
DB 58 -----ThrSerProLysTyrAspAlaGluLysLeuArgGlyAspArgLysGlnPro 74  
QY 589 -----ACAGCTCCGAGGGCAGCGCTGGAAGCGCTTCT 621  
DB 75 LeuTrpMetHisArgSerLeuMetArgLysSerGluArgProSerValTyrLeuAlaAla 94  
QY 622 GAGACCA-----ACGCGCCCGCAGTCTGCTTCACAGCTCTGAGGAGACGA 669  
DB 95 ArgSerArgHisProGlnLysGluThrProProSerGlnGluAspAlaLysGlnAlaAla 114  
QY 670 CCAGCCTCTTCTGAAAGGTGAAGGAGGCGGATGACACGATGACACCTCCGATAGTGAC 729  
DB 115 LysProSerSerProLysValLysLysSerLysGluAspLysAspLysSerAspSerGlu 134  
QY 730 ACGGATGGCTGACCTTGAAGAGCTTCAGATCGCCTTCGAGGAGCGGAACAGGAG 789  
DB 135 AlaGluSerIleValSerLysLysLysProArgLysLeuSerAlaLysGluLys 154  
QY 790 CCACACTGAGAG-----CCCTGAAAGGGATCCAGACT 822

DB 155 ProAspGluLysLysAspLeuLysGluArgLysAspSerLysLysGlyLysGluSer 174  
QY 823 -----CGCTG 828  
DB 175 AlaThrGluSerGluAspGluLysAlaGluLysGlyAlaLysLysAspArgLys 194  
QY 829 CGGAAGAACCGCGGAGGAGGTCGCCCGAGACTGTGGCTCCGAGCGCAGTACACT 888  
DB 195 GlySerLysLysGlyLysGluThrProSerAspSer---GlySerGluLysGlyAspAla 213  
QY 889 GTGAGGGCGCTCTGCGCCAGTAAGCAGGAGCCCGAG-----AACGAT 930  
DB 214 LysLysAspSerLysLysSerLysLysAspSerLysGlyLysGluSerAlaThrGluSer 233  
QY 931 CAGGGGTGTGTCCTCCAGCTGGGAAGATCACAGAGAGAGTAAG----- 975  
DB 234 GluGlyGluLysGlyAspAlaLysLysAspLysLysGlyLysGlySerLysLys 253  
QY 976 -----TTGAGGGAAGGCGCTCAGGACATCAAGATCAGGAG 1014  
DB 254 GlyLysGluSerAlaThrGluSerGluGlyLysGlyAspAlaLysLysAspLys 273  
QY 1015 CCTGAGAGACTTGGCGCGACCGAAG-----CCTGAATGTGAGGAGTTACGAC 1059  
DB 274 LysGlyLysLysGlySerLysLysGlyLysGluSerAlaThrGluSerGluGlyLys 293  
QY 1060 CCCAAGCGCTGTATTGCTATTGCGCGCAGCTCACACACAGGTTTATGATTTGCTGT 1119  
DB 294 GlyAspAlaLysLys----- 298  
QY 1120 GACCGCTGTGAAGAAATGTTTCATGGCGATGTGTGGGCATTTCTGAGGCTCGAGGAG 1179  
DB 299 -----AspAspLysLysGlyLys 304  
QY 1180 CTTTGGAAAGAAATGGGAAGACATATATCTGCCCCAACTGCACTTCTGCAAGTGAG 1239  
DB 305 LysGlySerLysLysGlyLysGlu-----SerAlaThrGlu 316  
QY 1240 GATGAGCTCATTCAGAAACGCGAGATCAGCAGGAAGCTAAATGAGACCTCGAGATGCT 1299  
DB 317 SerGluGlyLysGlyAspAlaLysLysAspLysLysGlyLysGlySerLys 336  
QY 1300 GATGCAACCGATTGTACAGTATAGAAACAATAGACGAGAGTCTAGCGAAGACCAAGG 1359  
DB 337 LysGly-----LysGluSerAlaThrGluSerGluGly 347  
QY 1360 ATAAGGCTAGATTGCAAAAGCTCAATCCAGTGGCAGCAAGAACTCAAGATCTTC 1419  
DB 348 GluLysGlyAspAlaLysLys---AspAspLysLysGlyLysLys----- 361  
QY 1420 CAGCCTGTGATAGAGCGCTGCTCTCAAAATGTTATTTGGCCCGGGTGTCTCACGCTG 1479  
DB 362 -----GlySerLysLys-----GlyLysGluSerAspSerLys 372  
QY 1480 GCGCAGCCGAGCTCGGTGTACTGCGAGTAATGACTGTATCTCTCAACACCGCCGAGGACA 1539  
DB 373 AlaGluGlyAspLysGlyAspAlaLysLysAsp---AspLysLysAspLysLysGlySer 391  
QY 1540 ATGAGTTTCTAAGCTCAGGTAAAGAACAGACAGAACCCCTAAACAAAGATGAAGATG 1599  
DB 392 LysLysGlyLysGluSerAlaThrGluSerGluGlyLysLysAspSerLysLysAsp 411  
QY 1600 AAGCAGAGAGCCCGAGTCTTCCGAAATGCGGTGCTCAGGAGGATTTAAATCTCTTCT 1659  
DB 412 LysAlaGlyLysAspProThrLysAlaGlyLysGlyAsp-----GluSer 428  
QY 1660 GTGCACAGAGACGAGCTCCAGAAAAAAGAGACACAGTCAAGACGAGCTGTGTGTC 1719  
DB 429 LysAspLysLysAspAlaLysLysAspSerLysLysGluLysLysAspGluLysLys 448  
QY 1720 CTTGCGCGAGTGTAGACACTCGGGAAGGAAGCAGCTGTGTGAGAGCAGCACCGCTGTGG 1779  
DB ----- 1779

Db 449 ProGlyGluAlaGluSerGluProLysAspSerAlaLysLysAspAlaLysLysAspAla 468

QY 1780 GCGAGCATCAATTACAGTCAGTAAGCCAGAAAGACTGCTGCT 1827

Db 469 LysLysAspAlaLysLysAspAlaLysLysAspAlaLysLysAspAla 484

RESULT 7

TRDN\_RABIT

ID TRDN\_RABIT STANDARD; PRT; 705 AA.

AC Q28820; Q28637; Q28643;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DE Triadin.

GM TRDN.

OS Oryctolagus cuniculus (Rabbit).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.

OX NCBI\_TaxID=9986;

EN [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORMS SKELETAL).

RC STRAIN=New Zealand white; TISSUE=skeletal muscle;

RX MEDLINE=93286104; PubMed=7685347;

RA Knudson C.M., Srang K.K., Moomaw C.R., Slaughter C.A., Campbell K.P.;

RT "Primary structure and topological analysis of a skeletal muscle-

RT specific functional sarcoplasmic reticulum glycoprotein (triadin).";

RL J. Biol. Chem. 268:12646-12654(1993).

RN [2]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORMS SKELETAL).

RC TISSUE=skeletal muscle;

RX MEDLINE=94298946; PubMed=8026576;

RA Peng M., Fan H., Kirley T.L., Caswell A.H., Schwartz A.;

RT "Structural diversity of triadin in skeletal muscle and evidence of

RT its existence in heart.;"

RL FEBS Lett. 348:17-20(1994).

RN [3]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORMS CARDIAC).

RC TISSUE=Heart muscle;

RX MEDLINE=96132942; PubMed=8550602;

RA Guo W., Jorgensen A.O., Jones L.R., Campbell K.P.;

RT "Biochemical characterization and molecular cloning of cardiac

RT triadin.;"

RL J. Biol. Chem. 271:458-465(1996).

RN [4]

RP CARBOHYDRATE-LINKAGE SITES, AND INTERCHAIN DISULFIDE BONDS.

RX MEDLINE=9606664; PubMed=7578102;

RA Fan H., Brandt N.R., Caswell A.H.;

RT "Disulfide bonds, N-glycosylation and transmembrane topology of

RT skeletal muscle triadin.;"

RL Biochemistry 34:14902-14908(1995).

CC -1- FUNCTION: May be involved in anchoring calsequestrin to the

CC junctional sarcoplasmic reticulum and allowing its functional

CC coupling with the ryanodine receptor.

CC -1- SUBUNIT: Homooligomer of variable subunit number; disulfide-

CC linked.

CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Sarcoplasmic

CC reticulum.

CC -1- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=6;

CC Comment=Additional isoforms seem to exist;

CC Name=Skeletal 1; Synonyms=ST1;

CC IsoId=Q28820-1; Sequence=ST1;

CC Name=Cardiac 1; Synonyms=CT1;

CC IsoId=Q28820-2; Sequence=VSP\_004458, VSP\_004460;

CC Name=Cardiac 2; Synonyms=CT2;

CC IsoId=Q28820-3; Sequence=VSP\_004458, VSP\_004461;

CC Name=Cardiac 3; Synonyms=CT3;

CC IsoId=Q28820-4; Sequence=VSP\_004466;

CC Name=Skeletal 2; Synonyms=ST2;

CC IsoId=Q28820-5; Sequence=VSP\_004462, VSP\_004464,

CC Name=Skeletal 3; Synonyms=ST3;

CC IsoId=Q28820-6; Sequence=VSP\_004464, VSP\_004465;

CC -1- TISSUE SPECIFICITY: Skeletal and cardiac muscle.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC

DR EMBL; U31540; AAC48496.1; -

DR EMBL; U31065; AAA31488.1; -

DR EMBL; U31555; AAC48497.1; -

DR EMBL; U34201; AAC48498.1; -

DR PIR; A45990; A45990.

KW Transmembrane; Sarcoplasmic reticulum; Glycoprotein;

FT INIT MET 0

FT DOMAIN 1 46

FT TRANSMEM 47 67

FT DOMAIN 68 705

FT CARBOHYD 74 74

FT VARSPLIC 624 624

FT VARSPLIC 264 285

FT VARSPLIC 264 307

FT VARSPLIC 286 705

FT VARSPLIC 308 705

FT VARSPLIC 415 415

FT VARSPLIC 416 424

FT VARSPLIC 584 584

FT VARSPLIC 585 591

FT VARSPLIC 645 705

SQ SEQUENCE 705 AA; 79003 MW; 13APID8475A1361 CRC64;

Alignment Scores:

Pred. No.: 0.00189

Score: 189.50

Percent Similarity: 37.18%

Best Local Similarity: 22.86%

Query Match: 4.04%

DB: 1

US-09-787-016a-1 (1-2610) x TRDN\_RABIT (1-705)

QY 697 GCGATACACGATGACCTCGATGACGAGGATGCTGACCTTGAAGAG--- 753

Db 117 GlyAspGluGluAspAspGluGlyAspGluAspThrAlaLysGlyGluIleGluGluPro 136

QY 754 ---CTTCAGATCGCTTCGAGAGCGGAGACAGGACCCACTGAGAGGCCCTTGAAG 810

Db 137 ProLeuLysArgLysAspIleHisLysGluLysGluLysGluGluGluArg 156

QY 811 GGGATCAGAGTGGCTG-----CGAAGACGCGGAGGAGGCTCCGCCGAG 861

Db 157 LysIleProThrLysValValHisLysGluLysGluLysGluLysValLysGlu 176

QY 862 ACTGTGGCTCCGAGGCCAGTGTGAGGGGCTCTCTGCCAGTAGACAGGAGGCC 921









QY 292 GAGGAGGACCTTAGCCCATCAACCCACCAAGAGATTCAGGANAACATGGGGTTT 351  
DB 93 GluGluGluAupLysAlaGlnValProThrGlnLys----- 104  
QY 352 CGAAGGACCATCTATCCGACAGCA-----GAGGGCGAGGAGCGGAGGCT 399  
DB 105 ---AlaAlaAlaProAlaAlaLysArgAlaSerLeuProGlnHisAlaGlyLysAlaAla 123  
QY 400 GACCCACTGGACGCCACCCACACAGCAGCAGCTGGGCGCTCTCCCTCGCGCGCAGTGGG 459  
DB 124 LysAlaSerGluSerSerSerGluGlu-----SerSerGluGluGlu 139  
QY 460 AGGAGCCACAGCAGCAGCTAGCGCGGTGAGAGAGTCTCGACCATTCGCGCGCGCGCGC 519  
DB 140 GluLysAlaLysLysLysProValGlnGlnLysAlaValLysProGlnAlaLysAla 159  
QY 520 AGGAGGAGCATCCCTGCTCTCCCTGAGGATTCGTGAGCCAGCCAGCTCTGCGCCGCCACA 579  
DB 160 ValArgProProLysLysAlaGlu---SerSerGluSerGluSerSerSer 178  
QY 580 GACGCGGACAGCCTCGAGGCGCGGTGAGAGCGCTTGTGAGACAGAGCGCGCC 639  
DB 179 GluAspGluAlaProGlnThrGlnLysProLysAlaAlaAlaThrAlaLysAlaPro 198  
QY 640 CAGTCTGCTTCCACAGCTGTGAAGCAACGACAGCCTCTTCTGAAAGGTG----- 690  
DB 199 ThrLysAlaGlnThrLysAlaProAlaLysProGlyProProAlaLysAlaGlnProLys 218  
QY 691 -----AAGGAGGGGATGACCAACAGATGACCTCCGATAGTACAGAGATGCG 738  
DB 219 AlaAlaAsnGlyLysAlaGlySerSerSerSerSerSerSerSerSerSer 238  
QY 739 CTGACCTTGAAGAGCTTCAGATCGCTTCCAGAGCGGGAACAGAGCCACTGAG 798  
DB 239 -----SerGluGluGluGluLysLysAlaAla 246  
QY 799 AGGCCCCCTGAAGGATCCAGAGTCGCTCGGAGAGAGCGCGGAGGAGGTCGCC 858  
DB 247 AlaProLeuLys-----LysThrAlaProLys 255  
QY 859 GAGACTGTGGCTCCGAGCCAGTACACTGTGAGGGGCTCTCCCGCAGTAAGCAGAG 918  
DB 256 LysGlnValValAlaLysAlaProValLysValThrAlaAlaProThrGlnLysSerSer 275  
QY 919 CCCGAGACGATCAGGGGCTGTGTCGCCAGGCTGGGAAAGATGACAGAGAGATGATG 978  
DB 276 SerSerGluAsp-----SerSerSerGluGluGluGluGlnLysLysPro 291  
QY 979 GAGGAAAGCGGCTCAGGACATCAAGATGAGGAGCTTGGAGACTTGGCGCGCAGCGAG 1038  
DB 292 MetLysLysLysAlaGlyProTyrSerSerValProProProSerValSerLeuSerLys 311  
QY 1039 CTGATGTGAGGGTTACGACCCCAAGCCCTGTATGCTATTTGCCGCGCAGCTCACAAC 1098  
DB 312 LysSerValGlyAlaGlnSerProLysLysAla---AlaAlaGlnThrGlnProAlaAsp 330  
QY 1099 AACAGGTTTATGTTGTGTGACCGCTGTGAGAAATGTTTCATGGCGAATGTGTGGCG 1158  
DB 331 Ser-----SerAlaAspSerSerGluGlu----- 338  
QY 1159 ATTCTGAGGCTCGAGGAGGCTTTGGAAAGCAATGGGGAAGACTATATCTGCCCAAC 1218  
DB 339 ---SerAspSerSerSerGlu-----GluGluLysLysThrProAla 351  
QY 1219 TGCACCATTTCTGCACTGAGGATGAGACTTATTCAGAAACGCGCATCAGCAGGAGCT 1278  
DB 352 LysThrValVal-----SerLysThrProAlaLysProAlaProValLysLysLys 368  
QY 1279 AATGAGACCTGGAGATCGATCGACCGGATTTACAGTATAGGACATAGGAGCAG 1338  
DB 369 AlaGluSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 379

QY 1339 AAGTCTAGCGAAGCAAGCGGATAAAGGCTAGTAATTCAGAAAGCTGCATAATCAAGTGGC 1398  
DB 380 ---SerSerGluAsp-----GluAlaProAlaLysProValSer 391  
QY 1399 AAGAAAGAACTCAAGATCTTCAGCCT---GTGATAGAGCGGCTGTGCTCAAAATGT 1455  
DB 392 AlaThrLysSerProLeuSerLysProAlaValThrProLysProProAlaAlaLysAla 411  
QY 1456 ATTGGCCCCGGTCTCTCAGTGGCGCAGCCGCTCGGTACTCGACGTAATGACTGT 1515  
DB 412 -----ValAlaThrProLysGlnProAlaGlySerGlyGln--L 424  
QY 1516 ATCTCTAAACACGCGCGCAGCAAT-----GAAGTTTCTAAGCTCAGTAA 1563  
DB 424 ySPGlnSerArgLysAlaAspSerSerSerSerGluGluSer--SerSerSerGlu 443  
QY 1564 GAACAAAGCCCAAG-----CTTAAAGAAAGATCAAGATCAACCCAG 1608  
DB 444 GluGluAlaThrLysLysSerValThrProLysAlaArgValThrAlaLys---Ala 462  
QY 1609 AAGCCAGTCTTCG---AAATGGGTGCTCAGCAGGTATTAATACTCTTCTGTGCAC 1665  
DB 463 AlaProSerLeuProAlaLysGlnAlaProArgAlaGlyLysAspSerSerSer 482  
QY 1666 AAGAGACGAGCTCCAGAAAGAACACAGTGAAGAGGAGGCTGTGCTCCTCG 1725  
DB 483 GluSerSerSerSerGluGluLysLysThrProLysPro-----ProAla 499  
QY 1726 CGGAGTCAACACTCGG-----AAGGAGCAGCTGT 1758  
DB 500 LysLysLysAlaAlaGlyAlaAlaValProLysProThrProValLysLysAlaAlaAla 519  
QY 1759 GAGAGCAGCAGCGCTGCTGGCGAGCGATCACAAT-----TACAATGCAAGTAAAGCA 1812  
DB 520 GluSerSerSerSerSerSerSerSerSerSerGluAspSerGluGluLysLysPro 539  
QY 1813 GAAAGACTGCTGCTCCTCGCG 1836  
DB 540 LysSerLysAlaThrProLysPro 547  
RESULT 10  
NP14\_HUMAN STANDARD; PRT; 699 AA.  
ID Q14978; Q15030;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Nuclear phosphoprotein p130 (Nucleolar 130 kDa protein) (140 kDa  
DE nucleolar phosphoprotein) (Nopp140) (Nucleolar and coiled-body  
DE phosphoprotein).  
GN NOLCL OR KIAA0035.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM ALPHA).  
RC TISSUE=Leukemia;  
RX MBLIN=9538590; PubMed=7657714;  
RA Pai C.-Y., Chen H.-K., Sheu H.-L., Yeh N.-H.;  
RT "Cell-cycle-dependent alterations of a highly phosphorylated nucleolar  
RL J. Cell Sci. 108:1911-1920(1995)."  
RN [2]  
RP SEQUENCE OF 3-699 FROM N.A. (ISOFORM BETA).  
RC TISSUE=Bone marrow;  
RX MBLIN=96051387; PubMed=7584026;  
RA Nomura N., Miyajima N., Sazuka T., Tanaka A., Kawarabayashi Y.,  
RA Sato S., Nagase T., Seki N., Ishikawa K.-I., Tabata S.;  
RT "Prediction of the coding sequences of unidentified human genes. I.  
RT The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by  
RT analysis of randomly sampled cDNA clones from human immature myeloid  
RT cell line KG-1.";

RL DNA Res. 1:27-35(1994).  
 RP [3]  
 RP ALTERNATIVE SPLICING  
 RA MEDLINE=96205319; PubMed=8630004;  
 RX Pai C.-Y.; Yeh N.-H.;  
 RT "Cell proliferation-dependent expression of two isoforms of the  
 RT nucleolar phosphoprotein p130.";  
 RL Biochem. Biophys. Res. Commun. 221:581-587(1996).  
 RN [4]  
 RP CHARACTERIZATION  
 RP MEDLINE=97168979; PubMed=9016786;  
 RX Chen H.-K.; Yeh N.-H.;  
 RA "The nucleolar phosphoprotein p130 is a GTPase/ATPase with intrinsic  
 RT property to form large complexes triggered by F- and Mg2+.";  
 RL Biochem. Biophys. Res. Commun. 230:370-375(1997).  
 RN [5]  
 RP CHARACTERIZATION  
 RP MEDLINE=20036810; PubMed=10567578;  
 RX Chen H.-K.; Pai C.-Y.; Huang J.-Y.; Yeh N.-H.;  
 RA "Human Nopp140, which interacts with RNA polymerase I: implications  
 RT for rRNA gene transcription and nucleolar structural organization.";  
 RL Mol. Cell. Biol. 19:8536-8546(1999).  
 CC -I- FUNCTION: Related to nucleogenesis, may play a role in the  
 CC maintenance of the fundamental structure of the fibrillar center  
 CC and dense fibrillar component in the nucleolus. It has intrinsic  
 CC GTPase and ATPase activities. May play an important role in  
 CC transcription catalyzed by RNA polymerase I.  
 CC -I- SUBUNIT: Interacts with RNA polymerase I 194 kDa subunit (RPAL94)  
 CC and with casein kinase-II.  
 CC -I- SUBCELLULAR LOCATION: Shuttles between the nucleolus and the  
 CC cytoplasm. At telophase it begins to assemble into granular-like  
 CC pre-nucleolar bodies which are subsequently relocated to nucleoli  
 CC at the early G1-phase.  
 CC -I- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=Alpha;  
 CC IsoId=Q14978-1; Sequence=displayed;  
 CC Name=Beta;  
 CC IsoId=Q14978-2; Sequence=VSP\_004338;  
 CC -I- PWM: Undergoes rapid and massive phosphorylation/dephosphorylation  
 CC cycles on CK2 and PKC sites. There is evidence suggesting that  
 CC CDC2 kinase phosphorylates p130 at the M-phase.  
 CC -I- SIMILARITY: Contains 1 Lish domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; Z34289; CAA84063.1; -;  
 DR EMBL; D21262; BAA04803.1; -;  
 DR PIR; I38073; I38073.  
 DR Genew; HGNC:15608; NOLC1.  
 DR GK; Q14978; -;  
 DR MIM; 602394; -;  
 DR GO; GO:0005737; C:cytoplasm; TAS.  
 DR GO; GO:0005730; C:nucleolus; TAS.  
 DR GO; GO:0005730; C:nucleolus; TAS.  
 DR GO; GO:0007049; P:cell cycle; TAS.  
 DR GO; GO:0007067; P:mitosis; TAS.  
 DR GO; GO:0006364; P:rRNA processing; TAS.  
 DR InterPro; IPR006594; Lish.  
 DR InterPro; IPR007718; SRP40\_C.  
 DR Pfam; PF05022; SRP40\_C\_1.  
 DR SMART; SM00667; Lish\_1.  
 DR PROSITE; PS50896; Lish; 1.  
 KW Nuclear protein; Phosphorylation; Repeat; GTP-binding; ATP-binding;  
 KW Alternative splicing.  
 FT DOMAIN 10 42  
 FT 11 X 12 AA APPROXIMATE REPEATS OF AN  
 FT LISH.  
 FT 84 566  
 FT DOMAIN ACIDIC SERINE CLUSTER.

FT REPEAT 84 95 ACIDIC SERINE CLUSTER 1.  
 FT REPEAT 125 136 ACIDIC SERINE CLUSTER 2.  
 FT REPEAT 167 178 ACIDIC SERINE CLUSTER 3.  
 FT REPEAT 221 232 ACIDIC SERINE CLUSTER 4.  
 FT REPEAT 264 275 ACIDIC SERINE CLUSTER 5.  
 FT REPEAT 325 336 ACIDIC SERINE CLUSTER 6.  
 FT REPEAT 363 375 ACIDIC SERINE CLUSTER 7.  
 FT REPEAT 425 436 ACIDIC SERINE CLUSTER 8.  
 FT REPEAT 470 481 ACIDIC SERINE CLUSTER 9.  
 FT REPEAT 519 529 ACIDIC SERINE CLUSTER 10.  
 FT REPEAT 555 566 ACIDIC SERINE CLUSTER 11.  
 FT DOMAIN 68 82 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 204 382 INTERACTS WITH RPA194.  
 FT DOMAIN 384 587 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 601 617 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT MOD\_RES 563 563 PHOSPHORYLATION (BY CK2) (BY SIMILARITY).  
 FT VARSPLIC 241 241 K -> KWTITSVRAE (in isoform Beta).  
 FT CONFLICT 3 3 /FTID=VSP\_004338.  
 FT CONFLICT 133 133 D -> A (IN REF. 2).  
 FT CONFLICT 231 231 R -> S (IN REF. 2).  
 FT CONFLICT 456 456 VA -> SV (IN REF. 2).  
 FT CONFLICT 456 456 S -> P (IN REF. 2).  
 SQ SEQUENCE 699 AA; 73720 MW; DFD4AD94EDF659FB CRC64;  
  
 Alignment Scores:  
 Pred. No.: 0.00497 Length: 699  
 Score: 182.00 Matches: 117  
 Percent Similarity: 36.10% Conservative: 83  
 Best Local Similarity: 21.12% Mismatches: 213  
 Query Match: 3.88% Indels: 141  
 DB: 1 Gaps: 19  
  
 US-09-787-016A-1 (1-2610) x NP14\_HUMAN (1-699)  
 QY 268 GAGCAAAAGGCGCCGAGCAATGAGGAGCGACCTAAGCGCATCAAAACCCACGACAA 327  
 Db 136 AspAspGluGluAspGlnLysGlnProValGlnLysGlnLysProGlnAlaLys 155  
 QY 328 GAGTTCAGGAAACATGCGGGTTTCGAGGACCATTCGCCAGCGAGGCGCGAGG 387  
 Db 156 AlaAlaLysAlaPro-----ProLysLysAlaLysSerSer 167  
 QY 388 GAGCGGAGGCTGAC---CCACTGGAGCGCCACCCACAGCAGCAGCTGGCGCTGTC 444  
 Db 168 AspSerAspSerAspSerSerSerGluAspGluProLysLysAsnGln----- 183  
 QY 445 CTGCGCGCAGTGGGAGGCGCCCAAGCGCAGCTGAGCGCGTGGAGCGATTCTCTGACCAT 504  
 Db 184 -----LysProLysIleThr----- 188  
 QY 505 GCGCGGCGCGCGGCGAGGAGCATGCTCTCCCTGGAGGATTCTGCTGAGCCCGACG 564  
 Db 189 -----ProValThrValLysAlaGlnThrLysAlaPro 199  
 QY 565 TCCTGCGCGCCACAGACGCGGAGCAGCGCTCCGAGGCGAGCGCTTGTAG 624  
 Db 200 ProLysProAlaAlaAlaProLysIleAlaAsnGlyLysAlaAlaSerSerSer 219  
 QY 625 ACCAGAGCGCGCCCGAGTCTGCTTCACAGCTGTGAGGAGGAGCAGCGCTCTCTGAA 684  
 Db 220 SerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 239  
 QY 685 AAG-----GTGAAAGGAGGGGATGAC 705  
 Db 240 LysLysThrValProLysLysGlnValValAlaLysAlaProValLysAlaAlaThrThr 259  
 QY 706 CACGATGACACCTCCGATAGTACAGCGATGGCTGACCTTGAAGAGCTTCAGAAATCG 765  
 Db 260 ProThrArgLysSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 279  
 QY 766 CTTCGAGGAGCGGCGACAGGAGCCC-----ACTGAGAGG 801  
 Db 280 LysProMetLysAsnLysProGlyProTyrSerTyrAlaProProProSerAlaProPro 299

```

QY 802 CCCCTCAAGAGGATCCAGAGTCCCTGCGGAAGAGCGCGGAGGAGGTCCTCCGCCGAG 861
D 803 ||||| : : : : : ||||| : : : : :
D 300 ProLysLysSerLeuGlyThrGlnProProLysLysAlaValGluLysGlnProVal 319
QY 862 ACTGTGGGCTCCAGGCGAGTGCACACTGTGGAGGGCTCTCCAGTAAAGCAGAGCC 921
D 863 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 320 GluSerSerGluAspSerSerAspGluSerSerSerSerSerGluGluGluLysPro 339
QY 922 GAGAACCATCAGCGGGTGTGTCCAGGCT : : : : : : : : : : : : : : : : : : : : :
D 923 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 340 ProThr---LysAlaValValSerLysAlaThrThrLysProProAlaLysLysAla 358
QY 964 AGAGACAGTAATGTGGAGAAAGCGGCTCAGGACATCAAGATCAGAG : : : : : 1014
D 965 ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 359 AlaGluSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 378
QY 1015 -----CTGGAGACTGGCGCCAGCGAAGCTGAATGTGAGGCTGACGCCCAACGCC 1068
D 1016 ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 379 LysProAlaGly-ThrThrLysAsnSerSerSerSerSerProAlaValThrThrLysSerPr 398
QY 1069 CTGTATTGCTTGGCGGAGCTCACAACAAGGTTTATGTTGTGTGACCGCTGT 1128
D 1070 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 398 AlaValLysProAlaAla-AlaProLysGlnProValGly----- 411
QY 1129 GAAGATGCTTTCATGCGGATGTGTGGCAATTCAGGCTGAGGAGCGCTTTGGAA 1188
D 1130 ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 412 -----GlyGlnLysLeuLeuThr 418
QY 1189 AGAATGGGGAAGACTATATCTGCCAACTGCACCATTCCTCAAGTCAGGATGAGCT 1248
D 1190 ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 419 ArgLysAlaAsp-----SerSerSerSerGluGlu 428
QY 1249 CATTGAGAAACCGCAGATCAGCAGGAGCTAA-----TGGAGACTGGA 1293
D 1250 ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 429 GluSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 447
QY 1294 GATGCTGATGGCAGCGATTGTCAAGTATA-----GGAAACAATA 1332
D 1295 ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 448 LysAlaThrAlaLysAlaAlaLeuSerLeuProAlaLysGlnAlaProGlnGlySerArg 467
QY 1333 GAGCAGAGTCTAGGAGACACGAGGATAAAGGTAGATTGAGAAGACTCCAAATCCA 1392
D 1334 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 468 AspSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 487
QY 1393 AGTGGCAAGAACTCAAGATCTTCCAGCTGTGATAGAGCGCTGGTCCCTCAAAA 1452
D 1394 ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 488 SerAlaValLysLys-----LysProGlnLysValAlaGlyGlyAlaAlaPro 503
QY 1453 TGTATTGGCCCGGGTGTGTCACTGGCGGCGAGCGAGCTCGGTGTACTGCAGTAATGAC 1512
D 1454 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 504 -----SerLysProAlaSerAla----- 509
QY 1513 TGTATCTCAACACCGCCGACGACATGAAGTTTCTAAGCTCAGGTAAAGACACAGAG 1572
D 1514 ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 510 -----LysLysGlyLysAlaGluSerSerSerSerSerSerSerSerSerSerSer 526
QY 1573 CCAAGACCTTAAGAAAGATGAAGTGAAG-----CCAGAGAACCCAGTCTTCGGAATGC 1629
D 1574 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 527 GluGluGluGluLysLysLysGlyLysGlySerProArgProGlnAlaProLysAla 546
QY 1630 GTGTCTCAGCAGGATTAAATCTCTCTGTGTCACAGAGCAGCTCCAGAAATAA 1689
D 1631 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 547 AsnGlyThrSerAlaLeuThrAlaGlnAsnGlyLysAlaAlaLysAsnSerGluGlu 566
QY 1690 GAGACACAGTGAAGAGCAGTGTGTGCTCCCTGGCGGAGTGAAGACTCGGGAG--- 1746
D 1691 ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 567 GluGluGluLysLysLysAlaAlaValValSerLysSerGlySerLeuLysLysArg 586
QY 1747 -----GAGCAGCTTGTGTGAGCAGCAGCGCG 1773
D 1748 ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 587 LysGlnAsnGluAlaLysGluAlaGluThrPro 598

```

```

RESULT 11
MYS3_HUMAN
ID MYS3_HUMAN STANDARD; PRT; 2004 AA.
AC Q92794;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DI 10-OCT-2003 (Rel. 42, Last annotation update)
DE MYST histone acetyltransferase 3 (Runt-related transcription factor
DE binding protein 2) (Monocytic leukemia zinc finger protein) (Zinc
DE finger protein 220).
GN MYST3 OR RUNXBP2 OR ZNF220 OR MOZ.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=96376968; PubMed=9782817;
RA Borrow J., Stanton V.P. Jr., Andresen J.M., Becher R., Behm F.G.,
RA Chaganti R.S.K., Clavin C.I., Distche C., Dube I., Frischauf A.M.,
RA Horman D., Mitelman P., Volinia S., Watmore A.E., Housman D.E.;
RA "The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses
RA a putative acetyltransferase to the CREB-binding protein.";
RL Nat. Genet. 14:33-41(1996).
CC -!- FUNCTION: May represent a chromatin-associated acetyltransferase.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- DISEASE: Involved in acute myeloid leukemias through a chromosomal
CC translocation t(8;16)(p11;p13) involving MYST3 and CREBBP.
CC -!- SIMILARITY: Contains 2 PHD-type zinc fingers.
CC -!- SIMILARITY: Belongs to the MYST (SAS/MOZ) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC SML; U47742; AAC50662.1; -.
CC Genew: HGNC:13013; MYST3.
CC
CC MIM; 601408; -.
CC GO; GO:0006323; P: DNA packaging; TAS.
CC InterPro; IPR005818; Histone_H1/H5.
CC InterPro; IPR003717; MOZ_SAS.
CC InterPro; IPR001965; Znf_PHD.
CC Pfam; PF01853; MOZ_SAS; 1.
CC Pfam; PF00628; PHD_2.
CC SMART; SM00526; H15; 1.
CC SMART; SM00249; PHD; 2.
CC PROSITE; PS01359; ZF_PHD_1; 1.
CC PROSITE; PS0016; ZF_PHD_2; 2.
CC Proto-oncogene; Chromosomal translocation; Zinc-finger; Repeat;
KW Nuclear protein.
FT ZN_FING 206 265 PHD-TYPE 1.
FT ZN_FING 259 313 PHD-TYPE 2.
FT DOMAIN 371 379 POLY-SER.
FT ZN_FING 538 560 C2HC-TYPE.
FT DOMAIN 788 801 POLY-GLU.
FT DOMAIN 989 995 POLY-GLU.
FT DOMAIN 1019 1026 POLY-ARG.
FT DOMAIN 1069 1078 POLY-GLU.
FT DOMAIN 1147 1150 POLY-LYS.
FT DOMAIN 1221 1242 GLU-RICH.
FT DOMAIN 1267 1302 GLU-RICH.
FT DOMAIN 1411 1414 POLY-GLU.
FT DOMAIN 1593 1597 POLY-SER.
FT DOMAIN 1643 1704 GLN/PRO-RICH.
FT DOMAIN 1897 1977 MET-RICH.
FT SITE 1546 1547 BREAKPOINT FOR TRANSLOCATION TO FORM
FT MOZ-CBB.
FT SEQUENCE 2004 AA; 225054 MW; 9FFBAC3792854BA CRC64;
SQ

```





Qy 1527 C-----GCCGCGAGCAATGAAGTTCTTAAGCTCAGGTAAAGAACAAGACCAAGCCAAAGCTTAA 1581  
Db 1634 rProGlnSerCysValValGluazgProproSerAsnGlnGlnGlnGlnInProrProPr 1654  
Qy 1584 AGAAAACATGAAGATGAAGCCAGCAGAGAGCCAGTCCTCCGAAATGGTGCTCAGGCAGG 1643  
Db 1654 oProProGlnGlnInProrProProProGln--ProGlnProAlaProGlnInPror 1673

Qy 1644 TATTAAAAATCTTCTTCGTGCAAGAGACCAGCTCCAGCAAAAAAAGAG 1692

Db 1673 oProProGlnGlnGlnInProrGlnGlnInProrGlnProGlnInProrGlnIn 1689

RESULT 12

ID CS58\_HUMAN STANDARD; PRT; 1012 AA.  
AC QBEVIO; Q9BMV4; Q9BXK3; Q9BZM3; Q9H421; Q9H4J6; Q9NZ22;  
DF 10-OCT-2003 {Rel. 42, Created}  
DT 10-OCT-2003 {Rel. 42, Last sequence update}  
DE Hepatocellular carcinoma-associated antigen 58 (Glioma-expressed  
antigen 2) [Transcription factor TZP] (Novel zinc finger protein).  
GN HCA58 OR GLRA2 OR TZP OR NZF OR C20ORF104.  
OS Homo sapiens (Human)  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Carnivora; Hominoidea; Homo.  
NCBI\_TaxID=9606;  
[1]  
RP SEQUENCE FROM N.A. (ISOFORM 2), TISSUE SPECIFICITY, AND DISEASE.  
RC TISSUE=Hepatoma;  
RX MEDLINE=20292308; PubMed=12097419;  
Wang Y., Han K.-J., Pang X.-W., Vaughan H.A., Qu W., Dong X.-Y.,  
Peng J.-R., Zhao H.-T., Rui J.-A., Leng X.-S., Burgess A.W.,  
Chen W.-P.;  
RA "Large scale identification of human hepatocellular carcinoma-  
associated antigens by autoantibodies.";  
RL J. Immunol. 169:1102-1109(2002).  
[2]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RN Cheng J.O., Kaneko S., Dan H.C., Testa J.R.;  
RT "Cloning and characterization of a novel transcription factor  
TZP".;  
RL Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.  
[3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21638749; PubMed=11780052;  
Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,  
Jones M., Pavlidis G., Almeida J.P., Babbage A.K., Baguley C.L.,  
Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M., Brown A.J.,  
Beusley O.P., Bird C.P., Blakey S.E., Bridgman A.M., Brown A.J.,  
Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,  
Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,  
Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,  
Coulson A., Coville G.J., Deadman R., Dharmaraj P., Dunn M.,  
Elliott A.G., Frankland J.A., Fraser A., French L., Garner P.,  
Graham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,  
Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,  
Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,  
Lehvasalo M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,  
Marsh V.L., Martin S.L., McConachie I.J., McLay K., McMurray A.A.,  
Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,  
Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,  
Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,  
Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showknee R., Sims S.,  
Stuice C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,  
Swann R.M., Sycomore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
Tracey A., Tromans A.C., Vaubin M., Wall M., Wallis J.M.,  
Whithead S.L., Whittaker P., Walley D.L., Williams L., Williams S.A.,  
Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
Rogers J.;  
RT "The DNA sequence and comparative analysis of human chromosome 20.";  
RL Nature 414:865-871(2001).  
[4]



QY	1066	GCCTGTATTCAGTTCGCGCAGAGCTCACAAACAGGTTTATGATTCGTGTCAGCGC	1125
Db	652	ValValArgCysIleCysGluValGlnGluAsnAspPheMetIleGlnCysGluGlu	671
QY	1126	TGTGAAGATGTTTCATGCGCATGTGTGCGCATTCCTGAGCTCGAGGAGGCTTTG	1185
Db	672	CysGlnCysTrpGlnHisGlyValCysMetGly	684
QY	1186	GAAAGGAAT---GGGAAGACTATATCTGCCAAACTGCACCATCTCTCAAGTCAGCAT	1242
Db	685	GluGluAsnValProGluLysTyrThrCysTyrValCys	697
QY	1243	GAGACTCATTCAGAAACGCGCAGATCGCAGGAGCTAAATGAGACCTGGAGATCTCAT	1302
Db	698	-----GlnAspProProGlyGlnArgProGlyPheLysTyr	709
QY	1303	GGCACCAGT-----TGACAAAGTATAGCAACAATACAGCAG	1338
Db	710	TrpTyrAspLysGluTrpLeuSerArgGlyHisMetHisGlyLeuAlaPheLeuGluGlu	729
QY	1339	AAGTCTAGCAAGAACCAAGGGATAAG-----GGTAGAATT	1374
Db	730	AsnTyrSerHisGlnAsnAlaLysLysIleValAlaThrHisGlnLeuLeuGlyAspVal	749
QY	1375	GAGAAAGCTGCAAT-----CCAAGTGGCAAGAACTCAAGATCTTCAGCCTGTG	1428
Db	750	GlnArgValIleGluValLeuHisGlyLeuGlnLeuLysMetSerIleLeuGlnSer	768
QY	1429	ATAGAGGCGCTGTGCTCCCTCAAAATGATGTCGCGCGGTGTGTGTCAGTCCGCGCAGCC	1488
Db	769	ArgGluHisProAspLeuProLeuTyrCysGlnProTyrLysGlnHisSerGlyGluGly	788
QY	1489	GACTCGGTACTGCACT-----ANTGACTGTATCTCAACACGCGCCGCGCACA	1539
Db	789	ArgSerHisPheArgAsnIleProValThrAspThrArgSerLysGluAlaProSer	808
QY	1540	ATGAAGTCTTCAAGCTCAGTAAAGAACAGACCAAGCCAAAGCCTAAAGAAAGATG	1599
Db	809	TyrArgThrLeu---AsnGlyAlaValGluLysProArgProLeu	822
QY	1600	AAGCCAGAGACCCAGCTTCGGAATCGGTGCTCAGGCGAGGTATTAATCTCTCT	1659
Db	823	-----AlaLeuProLeuProArgSerValGluGluSerTyrIleThrSer	837
QY	1660	GTGCACAGACACAGCTCCAGAAAAAAGAGACCAAGTGAAGAGCGAGTGTGTGTC	1719
Db	838	GluHisCysTyrGlnLysProArgAlaTyrTyrProAlaValGluGlnLysLeuValVal	857
QY	1720	CTGCGCGGAGTGAACATCTGGG-----AAGAACAGCTGTGTGTGAG	1761
Db	858	GluThrArgGlySerAlaLeuAspAspAlaValAsnProLeuHisGluAsnGlyAspAsp	877
QY	1762	AGCAGCACCG-----TCGTGGCGGAGCATCAATTAACAATGCA-----	1803
Db	878	SerLeuSerProArgLeuGlyTyrProLeuAspGlnAspArgSerLysGlyAspSerAsp	897
QY	1804	GTAAGCCCA-----GAAAGACTCTCTCTCCCTCGCGC	1836
Db	898	ProLysProGlySerProLysValLysGluTyrValSerLysLysAlaLeuProGluGlu	917
QY	1837	TCACTGTTGTAATAAGTATGATACCTAGGGGTGGCTCTGACCCCTCCGCTCT	1896
Db	918	AlaProAlaArgLysLeuLeuAspArgGlyGlyGluGlyLeuLeuSer---SerGlnHis	936
QY	1897	TTCTG 1902	
Db	937	GlnTip 938	
RESULT 13			
ID	TRDN	CANFA	
AC	P82179;	STANDARD;	PET; 700 AA.
DT	28-FEB-2003	(Rel. 41, Created)	

DT	28-FEB-2003	(Rel. 41, Last sequence update)	
DT	28-FEB-2003	(Rel. 41, Last annotation update)	
Db	Triadin.		
GN	Canis familiaris (Dog)		
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
CC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.		
CC	NCBI_TaxID=9615;		
OX	[1]		
RN	SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.		
RP	TISSUE=Heart, and Skeletal muscle;		
RC	MDLINE=99426545; PubMed=10497235;		
RA	Kobayashi Y.M., Jones L.R.;		
RT	"Identification of triadin 1 as the predominant triadin isoform expressed in mammalian myocardium.";		
RL	J. Biol. Chem. 274:28660-28668(1999).		
CC	IDENTIFICATION: May be involved in anchoring caldesmon to the junctional sarcoplasmic reticulum and allowing its functional coupling with the ryanodine receptor (By similarity).		
CC	FUNCTION: May be involved in anchoring caldesmon to the junctional sarcoplasmic reticulum and allowing its functional coupling with the ryanodine receptor (By similarity).		
CC	SUBCELLULAR LOCATION: Type II membrane protein. Sarcoplasmic reticulum.		
CC	ALTERNATIVE PRODUCTS:		
CC	Event=Alternative splicing; Named isoforms=3;		
CC	Name=Skeletal;		
CC	Isoid=P82179-1; Sequence=Displayed;		
CC	Name=Cardiac 1;		
CC	Isoid=P82179-2; Sequence=VSP_004001, VSP_004002;		
CC	Name=Cardiac 3;		
CC	Isoid=P82179-3; Sequence=VSP_004003, VSP_004004;		
CC	TISSUE SPECIFICITY: Skeletal and cardiac muscle.		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).		
CC	EMBL; AF165916; AAF00222.1; --		
DR	EMBL; AF165915; AAF00221.1; --		
DR	EMBL; AF165917; AAF00223.1; --		
KW	Transmembrane; Sarcoplasmic reticulum; Glycoprotein;		
KW	Alternative splicing.		
FT	INIT MET	0	
FT	DOMAIN	1 46	BY SIMILARITY.
FT	TRANSMEM	47 67	CYTOSOLASMIC (POTENTIAL).
FT	DOMAIN	68 700	POTENTIAL.
FT	CARBOHYD	74 74	LUMENAL (POTENTIAL).
FT	CARBOHYD	616 616	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	VARSPLIC	257 277	DOYAFCRYWDMFVHGDLPQ -> GKHSSEVAGSGKRTLG KKQIQ (in isoform Cardiac 1).
FT			/FTid=VSP_004001.
FT			Missing (in isoform Cardiac 1).
FT			/FTid=VSP_004002.
FT			E -> EPIKGKVKVPSLSKEKE (in isoform Cardiac 3).
FT			/FTid=VSP_004003.
FT			EKKVQKATKAKAIEKTVKPKAKAEHKEKESPTIKTD KPKPTSKETPSVETES -> GILQVVPVINCILFLVQFOQDE ELNVESKVRMIHVLHSPTSRSTSPILVISTCTCT (in isoform Cardiac 3).
FT			/FTid=VSP_004004.
FT			/FTid=VSP_004005.
SQ	SEQUENCE	700 AA; 78152 MW; P033E3A3A1BEE0C56 CRC64;	
Alignment Scores:			
Pred. No.:	0.0139	Length:	700
Score:	174.00	Matches:	107
Percent Similarity:	38.49%	Conservative:	77
Best Local Similarity:	22.38%	Mismatches:	190
Query Match:	3.71%	Indels:	104
DB:	1	Gaps:	20

[illegible]

Qy	1495	GTGTA	TCTGCAGTAAATGA	CTGTATCCTCAACAC-----GCCGACGACCAATGAAG	154
Dz	:	:	:	:	:
Db	413	GlnValLysAlaThrGluArgAlaGlyLeuThrSerAlaAlaSerThrLysLys	432		
Qy	1546	TTTCTAAGCTCAGGTAAAGAACAACAGAACCCAAAGCGCTAAAGAAAAGATGAAGATGAAGCCA	1605		
Dz	:	:	:	:	:
Db	433	AlaValProGlyLysLysGluGluLysThrThrLysThrValGluGlnGluIleArgLys	452		
Qy	1606	GAGAAGCCCCAGTCCTCCGAAATGGGTGCTCAGGCAGCGATTATAAATCTCTTCTTGTCAC	1665		
Dz	:	:	:	:	:
Db	453	GluLysSerGly-----LysThrSerThrAlaser	462		
Qy	1666	AAGAGACCAGCTCCAGAA--AAAAAGAGACACACAGTGAAGAGCCA-----GTGGTG	1716		
Dz	:	:	:	:	:
Db	463	LysAspLysGluProGluIleLysLysAspGluLysMetProLysAlaAspLysGluVal	482		
Qy	1717	GTCCCTCGCGGAGTGAAGCA---CTCGGAAGAAGCAGCTTGTGAGAGCAGCACCGCG	1773		
Dz	:	:	:	:	:
Db	483	LysProLysProProGlnInserGlnValLysLysGluLysSerGluSerGlnValLys	502		
Qy	1774	TCGTGGCGGAGCATCAATACAACTGAGTAAAGCCAGAAAAGACTGTGCT	1827		
Dz	:	:	:	:	:
Db	503	LysGluAlaLysProGluGln---AspIleAlaLysProGluLysThrValSer	519		
<hr/>					
RESULT 14					
RRE1 HUMAN					
ID	RRE1_HUMAN	STANDARD;	PRT;	755 AA.	
AC	Q92766;				
DT	15-JUL-1998 (Rel. 36, Created)				
DT	15-OCT-1998 (Rel. 36, last sequence update)				
DT	10-JUL-2003 (Rel. 42, last annotation update)				
DE	RAS-responsive element binding protein 1 (RREB-1).				
GN	RREB1.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Thyroid carcinoma;				
RX	MEDLINE=96413283; PubMed=8816445;				
RA	Thiagalingam A., de Bustros A., Borges M., Jasti R., Compton D.,				
RA	Diamond L., Mabry M., Ball D.W., Baylin S.B., Neikins B.D.;				
RT	"RREB-1, a novel zinc finger protein, is involved in the				
RT	differentiation response to Ras in human medullary thyroid				
RT	carcinomas.";				
RL	Mol. Cell. Biol. 16:5335-5345(1996).				
CC	-! FUNCTION: TRANSCRIPTION FACTOR THAT BINDS SPECIFICALLY TO THE				
CC	DISTAL RAS-RESPONSIVE ELEMENT [RRE] IN THE CALCITONIN GENE				
CC	PROMOTER AND AUGMENT THE RAS/RAF-MEDIATED TRANSCRIPTIONAL RESPONSE				
CC	OF THAT PROMOTER. MAY BE INVOLVED IN RAS/RAF-MEDIATED CELL				
CC	DIFFERENTIATION				
CC	-! SUBCELLULAR LOCATION: Nuclear.				
CC	-! TISSUE SPECIFICITY: EXPRESSED IN HEART, PLACENTA, LUNG, LIVER,				
CC	SKELETAL MUSCLE, KIDNEY AND PANCREAS. NOT FOUND IN THE BRAIN.				
<hr/>					
THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION					
between the Swiss Institute of Bioinformatics and the EMBL Outstation -					
the European Bioinformatics Institute. There are no restrictions on its					
use by non-profit institutions as long as its content is in no way					
modified and this statement is not removed. Usage by and for commercial					
entities requires a license agreement (See http://www.isb-sib.ch/announce/					
or send an email to license@sib-sib.ch).					
<hr/>					
EMBL;	U26914; AAB19094.1; --				
HSSP;	P08045; 1ZNH.				
DR	TRANSFAC; T01975; --				
DR	Genew; HGNC:10449; RREB1.				
DR	MIW; G02209; --				
DR	GO; G0007275; P:development; TAS.				
DR	GO; G0007265; P:RAS protein signal transduction; TAS.				
DR	GO; G0006366; P:transcription from Pol II promoter; TAS.				

DR InterPro; IPR007087; Znf C2H2.  
 DR Pfam; PF00096; Zf\_C2H2; 4.  
 DR SMART; SM00355; Znf C2H2; 4.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 4.  
 DR PROSITE; PS00157; ZINC\_FINGER\_C2H2\_2; 4.  
 KW Transcription regulation; Activator; zinc-finger; Metal-binding;  
 XW DNA-binding; Nuclear protein; Repeat.  
 FT ZN\_FING 315 337 C2H2-TYPE.  
 FT ZN\_FING 461 483 C2H2-TYPE.  
 FT ZN\_FING 580 602 C2H2-TYPE.  
 FT ZN\_FING 608 630 C2H2-TYPE.  
 SQ SEQUENCE 755 AA; 79865 MW; 28D863EF62FP8415 CRC64;

Alignment Scores:  
 Pred. No.: 0.016 Length: 755  
 Score: 173.00 Matches: 138  
 Percent Similarity: 33.66% Conservative: 70  
 Best Local Similarity: 22.33% Mismatches: 216  
 Query Match: 3.69% Indels: 194  
 DB: 1 Gaps: 30

US-09-787-016a-1 (1-2610) x RRE1\_HUMAN (1-755)

QY	305	AGCCATCAACCCACCAAG	-----AGTTCAGAAACATGGG	346
Db	37	GlnProSerProCysProAlaProGlyProSerLeuProValThreueGlyProSerGly	56	
QY	347	GTTCGAGGACCACTATCGCAAGCAGAGGGG	-----	382
Db	57	IleLeuGluSerProMetAlaProAlaProAlaThProGluProProAlaGlnPro	76	
QY	383	---CAGGACCGGAGGCTGACCCACTGGAGCGGCCACCCACAGCAGCAGCTGGGCC	439	
Db	77	LeuGlnGlyProValGlnLeu	-----	83
QY	440	TGTCCCTCGCGCGAGTGGGAGGAGCCCAAGCCACTGAGCGGAGAGTTCCTGA	499	
Db	84	-----AlaValProIleTyrSerSerAlaLeuValSerSerPro	96	
QY	500	CCATTGCGGCGCGCGCA	-----GGAGGAGCATGCTCTCCTCGGAGGATT	550
Db	97	ProLeuValGlySerSerAlaLeuLeuSerGlyThrAlaLeuLeuArgProLeuArg	115	
QY	551	CTGGTGAAGCCACGT	-----CCTGCCCCCGCCACAGACGCGGAGCAGCCT	595
Db	116	-----ProLysProProLeuLeuLeuProLysProValThreueGlyPro	132	
QY	596	CGAGGCGAGCTGGAAGCGTCTCAGACCA	---GAAGCGCGCCCGAGTCTGCTCCA	652
Db	133	ProLeuAlaSerIleAlaGlnIleIleSerSerValSerAlaProThrLeuLeuLys	152	
QY	653	CAGCTGTGAAGAACGACCCAGCT	-----	676
Db	153	ThrLysValAlaAspProGlyProAlaSerThrGlySerAsnThrThAlaSerAspSer	172	
QY	677	CTTCTGAAAGGTGAAGAGGGATGACACGATGACCTCCGATAGTACAGGATG	736	
Db	173	LeuGlyGlySerValProLysAlaAlaThrThAlaThrProAlaAlaThrThrSer	191	
QY	737	GCTGACCTTGAAGAGCTTCAGATCGCTTCGACGAGGAGCGGGAACAGGAGCCACTG	796	
Db	192	-----Pro-LysGluSerSerGluProPro	199	
QY	797	AGAGGCCCTGAAAGGATCCAGAG	-----TGGCTGCGGAAGAACGCCCGGAGGAGG	850
Db	200	---AlaProAlaSerSerProGluAlaAlaSerProThrGluGlnGlyProAlaArgTh	218	
QY	851	GTCCGCCGAGACTGTGGGCTCGAGGCCAGTACACTGTGGAGGCGCTCCGCCAGTA	910	
Db	218	zSerLysLysArgGlyArgLysArgGlyMet	-----ArgSerArgPro	232
QY	911	AGCAGGAGCCGAGAACGATCAGGGGGTGTGTGCC	-----	945

Db	233	-----ArgAlaAsnSerGlyGly-ValAspLeuAspSerSerGlyGluPheAla	249	
QY	946	-----	-----CAGCGTG	952
Db	249	erIleGluLysMetLeuAlaThrThrAspThrAsnLysPheSerProPheLeuGlnThrA	269	
QY	953	GGAAGATGACAGAGAGATTAAGTTGAGGGAAAGGGCGCTCAGGACATCAAGATGAGG	1012	
Db	269	laGluAspAsnThrGlnAspGluValAlaGly	-----A	280
QY	1013	AGCTGAGAGACTTGGCGCCGACCGAGCCTGAATGTGAGGGTTACGACCCC	-----AACGCC	1069
Db	280	laProAlaAspHisHisGlyProSerAspGluGlnGlySerProProGluAspLysL	300	
QY	1070	TGTATTGCATTTGCCGCGCAGCTCACAACACAGTTTATG	-----ATTGCTGACCGCT	1126
Db	300	euLeuArgAlaLysArgAsnSerTyrThrAsnCysLeuGlnLysIleThrCysProHisC	320	
QY	1127	GTGAAGAATGGTTT	-----CATGGCGATT	1150
Db	320	ysProArgValPheProTpaAlaSerSerLeuGlnArgHis**LeuThrHisThrAsps	340	
QY	1151	GTGGGGCATTTCTGAG	-----GCTGAGGGAGCGCTTTTGAAGAGGANTGGGAG	1201
Db	340	erGlnSerAspAlaGluThrAlaAlaAla**GlyGluValLeuAspLeuThrSerArgA	360	
QY	1202	AC	-----TATATCTGCCAACTGCACCA	1225
Db	360	spArgGluGlnProSerGluGlyAlaThrGluLeuArgGlnValAlaGlyAspAlaProV	380	
QY	1226	TTCTGCAAGTCAGAGATGAGATCATTCAGAAACGCGCAGATCAGCAGGAGAGCTAAA	1281	
Db	380	alGluGlnAlaThrAlaGluThrAlaSerProValHisArgGluGluHisGlyArgGlyG	400	
QY	1282	-----TGAGACCTCGAGATCTGATGGCAGCGATTTGTACAAGTATAGAAACAATAGAGC	1336	
Db	400	luSerHisGluProGluGluGluHisGlyThr	-----GluG	412
QY	1337	AGAACTCTAGCAAGACCAAGGATAAAGGTAGATTTGAGAAGCTGCAAAATCAAAGT	1396	
Db	412	luSerThrGlyAspAlaAspGlyGlyArgGlyArgValGluGln	-----ProGluP	429
QY	1397	GCAGAGAACTCAAGATCTTCAGCCTGTGATAGAGCGCGCTGCTCCTCAAAATGTA	1456	
Db	429	roGlyProGlyLeuArgHisGlnAlaHisGlyLeuGlnAlaGlyGlyGlyArgArgG	449	
QY	1457	TTGGCCCGGCTGTGTCACTGCGCAGCCCGACTCGCTGCTACTGCTAGTAACTACTGTA	1516	
Db	449	lyArgProGlyGlyAla	-----AlaSerGlnGlnLysLeuAlaCysaspThr	465
QY	1517	TCCTCAACACCGCCGCGACAAATGAATTTCTTAAGCTCAGGTAAAGAACAGAGCCAA	1576	
Db	466	-----CysGlyLysSerPheLysPheLeuGlyThrLeuSerArgHisArg	481	
QY	1577	AGCTTAAGAAAGATGAAGATGAAGCCAGAGAGCCCGCTTCCGAAATCGCGTGCTC	1636	
Db	481	ysAlaHisGlyArgGlnGluProLysaspGlyLysGlyAsp	-----	494
QY	1637	AGGCGAGTATTAAATCTTCTGTGCACAAGAGACAGCTCCAGCAAAAAA	1689	
Db	495	-----GlyAlaThrThrAlaGluGluGlyProSerProAlaProGluGlnGluLysP	513	
QY	1690	-----GAGACACAGTGAAGAGCGAGTGGTGGTCCCTGCGCGAGTGAAGCACTCGGGA	1744	
Db	513	roProGluThrProAlaGluValValGluSerAlaProGlyAlaGlyGluAlaProAlaG	533	
QY	1745	AGGAGCAGCTGTGAGAGCAGCAGCGCGTGG	-----TGGG	1780
Db	533	luLysLeuAlaGluGluThrGluGlyProSerAspGlyGluSerAlaAlaGluLysArgS	553	
QY	1781	CGAGCATCAATTCATTCAGTAAGCCAGAAAGACAGCTGCTGCC	1830	
Db	553	erSerGluLysSerAspAspLysLysPro	-----LysThrAspSerPro	568





QY	580	-----GACGCCGAGACAGCCTCCGAGGCGAGCGTGGAAAGCCGCTCTGAGACACAGA	630
DB	698	AlaGlnAlaGlnValThrAlaProProLeuLysThrAspSerAlaLysThrSerGln	717
QY	631	AGCGGCCCCAGTCTGCTCCACA-----CCTGTGAAGGAACGACCA	672
DB	718	SerPheProProThrGlyAspThrIleThrProLeuAspSerLysAlaMetProArgPro	737
QY	673	GCCTCTTCTGAAAGGTGAAAGGAGGGGATGACACAGCATGACACCTCCGATATAGTCAGCG	732
DB	738	AlaSerAspSerLysIleVal-----SerHisProGlyProThrSerGluSerLys	754
QY	733	GATGGCGTGCACCTTGAAGAGCTTCAGATCGCTTCGACGAGGCGGGAACAGGAGCC	792
DB	755	AspProValGlnLysLysGluProLysLysAlaGlnThrLysValThrProLysPro	774
QY	793	ACTGAGAGGCCCTG-----AAAGGGATC-----	816
DB	775	AspThrLysProValProLysGlySerProThrProSerGlyThrArgProThrThrGly	794
QY	817	-----CAGATGCGCTCGGAGAAAGCGCGGAGAGAGCGT-----	852
DB	795	GlnAlaThrProGlnSerGlnProProLysProProGlnSerArgArgPheSer	814
QY	853	-----CCCGCCGAGACT	864
DB	815	LeuAsnLeuGlyGlyIleAlaAspAlaProLysSerGlnProThrProGlnGluThr	834
QY	865	GTGGGC-----TCCGAGCGCAGTGCACCTGTG	891
DB	835	ValThrGlyLysLeuPheGlyPheGlyAlaSerIlePheSerGlnAlaSerAsnLeuIle	854
QY	892	GAG-----GGCGTCTGCCCCCAAGCAG	915
DB	855	SerThrAlaGlyGlnGlnAlaProHisProGlnThrClyProAlaAlaProSerLysGln	874
QY	916	GAGCCCGAGAACGATCAGGGGGTGTGTCACAGCT-----	951
DB	875	AlaProProSerGlnThrLeuAlaAlaGlnGlyProProLysSerThrGlyGlnHis	894
QY	952	-----CGGAAAGATGACAGAGAGAGT	972
DB	895	ProSerAlaProAlaLysThrAlaValLysLysGluThrLysGlyProAlaAlaGlu	914
QY	973	AAGTTGAGGGAAGCGCGTCTAG-----GACATCAAGATGAGGAG-----	1014
DB	915	AsnLeuGluAlaLysProAlaGlnAlaProThrValLysLysAlaGluLysAspLysLys	934
QY	1015	-----CCTGAGACTGGGC-----	1029
DB	935	HisProProGlyLysValSerLysProProThrGluProGluLysAlaValLeuAla	954
QY	1030	-----CGACCGAAGCCTGATGTGAGGTGTAGACCCCAAGCC	1068
DB	955	GlnLysProAspLysThrThrLysProLysProAlaCysPro-----	968
QY	1069	CTGTATTGCATTTCGCCG-----CAGCCTCACAACAAC	1101
DB	969	-----LeuCysAlaGThrGluLeuAsnValGlySerGlnAspProProAsnPheAsn	985
QY	1102	AGGTTTAGATTTCG-----TGTCACCGCTGT-----	1128
DB	986	ThrCysThrGluCysLysAsnGlnValCysAsnLeuCysGlyPheAsnProThrProHis	1005
QY	1129	-----GAAGATGTTTTCATGCGGATTTGTGCGCATTTCTGAGGCTCGAGG	1176
DB	1006	LeuThrGluIleGlnGluTrp-----	1012
QY	1177	AGGCTTTTGAAGGAATGGGGAAGACTATATCTGCCAAACTGCACCATTCGCAAGTG	1236
DB	1013	-----LeuLysLeuAsnCys-----GlnThr	1019

Qy	1237	CAGATGAGACTCA	TT	CAGAAACGGC	GAGATCAGCAGAGACTAAATGGGACCTGGAGAT	1239
Db	1020	GlnArgAlaileSer	GlyGlnLeuGlyAspMetAspLys	MetProAla	1036	
Qy	1297	GCTGATGGCACC	ATTGTTACAAGTATAGGAAACAATAGAGCAGAAAGTCTACGAGAACCAA	1356		
Db	1037	SerSerGlyProLysAlaSerProValProAlaProAlaGluProProGln	1054			
Qy	1357	GGGATAAGGGT	AGAAATGAGAAAGCTGCMAATCCAAAGTGGCAGAGAGAACTCAAGATC	1416		
Db	1055	-----LysThrProThrAlaAlaHisAlaLysGlyLysLysGluThrGlu	1070			
Qy	1417	TTCCAG	-----CCTGTGATAGAGCGCGCTGGTGCTCAAAATGT	1455		
Db	1071	ValLysAlaGluThrGluLysGlnLeuProGluLysGluThrProSerIleGluLys	1089			
Qy	1456	ATTGGCCCCGGGTGCTGCTCAGTGGCGCAGCCGACTCGGTGTA	CTGCAGTAATGACTGT	1515		
Db	1090	-----ThrProAlaVal	1098			
Qy	1516	ATCCTCAACACACGCCCGCAGCGACAATGAAGTTCTTAAGCTCAGGTAAAGAACAGAACCA	1575			
Db	1099	LysLeuGluGluSerGluValThrLysSerLeuValSerValLeuProGluLysLysPro	1118			
Qy	1576	AAGCCTAAAGAAAGATGAAGATCAAGCCAGCAAGCCCAAGTCTTCGAAATCGGTGCT	1635			
Db	1119	SerGluGluGluLysAlaLeuProAlaAspLysLysGluLysLysProProAlaAlaGlu	1138			
Qy	1636	CAGCAGGTATTA	AAATCTCTCTGTGCCAAGAGACCCTCCA	1680		
Db	1139	AlaProProLeuGlu	-----GluLysLysProIleProAspGlnLysLeu	1154		
Qy	1681	-----GAAAAAAGAGACCCAGGTGAAGAG	1707			
Db	1155	ProProAspAlaLysProSerAlaSerGluGluGluLysArgAspLeuLeuLysAla	1174			
Qy	1708	GCATGTGTGCTCC	TCCGCGGAGTGAAGCACTCGGAGCAACA	1761		
Db	1175	HisValGlnIleProGluGluGly	-----ProIleGlyLysValAlaSerLeuAlaCysGlu	1193		
Qy	1762	AGCAGCACCGCGT	CTCGCGCGAGCATCAATAATGCAGTAAAGCCCAAGAACT	1821		
Db	1194	Gly	-----GluGlnGlnProAspThrArgProGluAspLeu	1205		
Qy	1822	GCTGCTCCCTCGCGCTGCTGTGTATATAA	1851			
Db	1206	ProGlyAlaThrProGlnThrLeuProLys	1215			

Search completed: April 28, 2004, 10:55:23  
Job time : 70.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 28, 2004, 10:46:07 ; Search time 129.5 Seconds  
(without alignments)  
12718.185 Million cell updates/sec

Title: US-09-787-016A-1  
Perfect score: 4689  
Sequence: 1 crcggtggccgctcgccac.....actcttaagatcatatcctg 2610

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p,model -DEV=xlp  
-O=/cpn2.1/USPTO spool\_p/US09787016/runat 28042004 094329 26919/app query.fasta\_1.2759  
-DB=SPTRMBL\_25 -OFFT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09787016 @CEN 1.1 142 @runat 28042004 094329 26919 -NCFU=6 -ICFU=3  
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG  
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOB=6  
-FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTRMBL\_25:  
1: sp\_archaea:  
2: sp\_bacteria:  
3: sp\_fungi:  
4: sp\_human:  
5: sp\_invertebrate:  
6: sp\_mammal:  
7: sp\_mhc:  
8: sp\_organelle:  
9: sp\_phase:  
10: sp\_plant:  
11: sp\_prodent:  
12: sp\_virus:  
13: sp\_vertebrate:  
14: sp\_unclassified:  
15: sp\_rvirus:  
16: sp\_bacteriap:  
17: sp\_archaea:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	332	7.1	2016	5 Q9VG78	Q9VG78 drosophila

ALIGNMENTS	
RESULT 1	
Q9VG78	PRELIMINARY; PRT; 2016 AA.
ID	Q9VG78
AC	Q9VG78; 2016 AA (TREMELrel. 13, Created)
DF	01-MAY-2000 (TREMELrel. 13, Last sequence update)
DT	01-MAY-2000 (TREMELrel. 13, Last sequence update)
DE	CG6525 protein.
GN	SPP OR CG6525.
OS	Drosophila melanogaster (Fruit fly).
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC	Ephydroidea; Drosophilidae; Drosophila.
OX	NCBI TaxID=7227;
RN	[1]
RP	SEQUENCE FROM N.A.
KC	STRAIN=Berkeley;
RX	MEDLINE=20196006; PubMed=10731132;
RA	Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA	Sutton R.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA	Brandon R.C., Rogers Y.-H.C., Blazer E.G., Champe M., Pfeiffer B.D.,
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA	Abril J.F., Adayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,

Q9W352 drosophila  
Q9VUB5 drosophila  
Q8C3Y1 drosophila  
Q9W0T2 drosophila  
Q9VUB8 drosophila  
Q9W0T1 drosophila  
Q7Z7D6 homo sapien  
Q8I6F0 euplores cr  
Q7Z2C9 drosophila  
Q95YM2 procamburus  
Q8C9E9 mus musculus  
Q8C9E0 mus musculus  
Q7SZE4 brachydanio  
Q74508 schizosacch  
Q7SZX6 brachydanio  
Q8IWP2 homo sapien  
Q8IWP1 homo sapien  
Q8IWP0 homo sapien  
Q8IWN9 homo sapien  
Q8IWN8 homo sapien  
Q8IWN7 homo sapien  
Q86SQ1 homo sapien  
Q9BUV3 homo sapien  
Q95PU8 chironomus  
Q78891 drosophila  
Q9W596 drosophila  
Q35540 mus musculus  
Q28687 oryctolagus  
Q91255 petromyzon  
Q871Y7 neurospora  
Q83291 figu rubrip  
Q8VDN7 mus musculus  
Q9QX75 rattus norv  
Q9UQ40 homo sapien  
Q9HFW4 ustilago ma  
Q9UBA8 homo sapien  
Q9UQ35 homo sapien  
Q8H2W3 canis famil  
Q8IP58 drosophila  
Q9VNX6 drosophila  
Q99192 mus musculus  
P97868 mus musculus  
Q9EQJ5 mus musculus  
O45407 caenorhabdi

RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolekhov S.,  
RA Borkova D., Borchan M.R., Bouck J.P., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
de Pablos B., Delcher A., Deng X., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg M.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jaitai M., Kalush P., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.A., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reine B.C., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RA "The genome sequence of *Drosophila melanogaster*";  
RT Science 287:2185-2195(2000).  
DR EMBL: AEO03695; AAP54807.1; --  
DR FlyBase: FBgn0038041; Spp.  
DR GO: GO:0003677; F:DNA binding; IEA.  
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
DR InterPro: IPR006576; BRK.  
DR InterPro: IPR001965; Znf\_PHD.  
DR Pfam: PF00628; PHD; 1.  
DR SMART: SM00249; PHD; 1.  
DR SMART: SM00592; BRK; 1.  
DR PROSITE: PS01359; ZF\_PHD\_1; 1.  
DR PROSITE: PS50016; ZF\_PHD\_2; 1.  
DR SQ SEQUENCE 2016 AA; 221998 MW; F6B1A332BA710F5E CRC64;

Alignment Scores:  
Pred. No.: 1,47e-14 Length: 2016  
Score: 332.00 Matches: 159  
Percent Similarity: 32.52% Conservative: 68  
Best Local Similarity: 22.78% Mismatches: 221  
Query Match: 7.08% Indels: 250  
DB: 5 Caps: 27

US-09-787-016A-1 (1-2610) x Q9VG78 (1-2016)

QY 256 GGAGCAGGATGACGACCAAGCGGACCGGCAATGAGGAGCAGCTAAGGCCATCAAA 315  
Db 676 GlyValGlyAsnAlaAlaAlaAspProAsn-----ileMet 689  
QY 316 CCCACAGCAAGAGTTCAGAAACATGGGTTTCGAAAGGACCACTATCGCCAGCGA 375  
Db 690 AspThrAlaAsnGluAspGluIleThrAlaAspPheLeuGlnHisValValGlyLeuIle 709  
QY 376 GAGGCGCAGCGGACGCGAGGTGACCCACTGGAGCGCGCCACCCACACGACGAGCTG 435  
Db 710 GluGluAspLysGlnPheGluAla----- 717  
QY 436 GGCCTGTCCCTGGCGGCGAGTGGGAGGACGCCCAAGGCGACTGAGCGGTGGAGGATTC 495  
Db 718 -----GluValValLysGlnVal 723  
QY 496 CTGACCATTCGCGGCGCGCGCAGGAGGAGCATGCTCTCCCTGGAGGATTCGTGT 555  
Db 724 LeuAla----SerThrGluProGlyThrLeuAspAlaIleValSerMetProThrSerIle 742

QY 556 GAGCCCAAGTCTGCTGCCCGCCACAGACCGGACGAGCAGCTCCGAGGGCAGCTG---GAA 612  
Db 743 GluProValAspValProGlnAlaHisThrAsnLeuLeuProAsnAlaSerLeuThrGlu 762  
QY 613 AGCGTCTTCGACAGCAGAGCGGCCCCCAGCTCTGTTCCACA-----GCT 557  
Db 763 ProAlaGlnSerMetThrSerLeuProIleAlaCysSerThrProSerArgSerValAla 782  
QY 658 GTGAGGAACACGACGACCTCTTCTGAAAGGTGAAAGGAGGGGATGACCAACGATGACACC 717  
Db 793 AlaSerThrProProThrSerAlaLysValValArgGly----- 795  
QY 718 TCCGATAGTGCAGCGATGGCTGCCTTGAAGAGGCTTCAGAAATCGCTTCGCGAGGAG 777  
Db 796 -----TyrglyArgVal 799  
QY 778 CGGGAACAGGACCCACTGAGAGGCCCTCTGAAAGGATCCAGAGTCCGCTCGGGAAGAG 837  
Db 800 IleTyLeuProProIleGluAlaPro-----ThrThrArgAlaLys 813  
QY 838 CGCGCGGAGGAG-----GGTCCCGCGGAGACTGTGGGTCTCCGAGGCC--- 879  
Db 814 ArgArgAlaGlnPheProSerAlaProGlyMetAlaAlaThrSerSerSerAspAlaGly 833  
QY 879 ----- 879  
Db 834 AsnLeuSerPheGlyGluSerSerLeuAspAlaSerIleAsnGlnProLeuAsnThrSer 853  
QY 880 -----AGTGACACTGTGGAGGCGCTCTGCCCACTAAGCAGGAGCCCGAGAAAGAT 930  
Db 854 SerLeuSerAsnAspSerGlnProGlySerGlyProLysArgProAsnPro---ArgGlu 872  
QY 931 CAGGGGGTGTGTCCCGAGGCTGGGAAAGATGACAGAGAGTAAAGTGTGAGGAAAGGCG 990  
Db 873 ProSerMetAlaArgArgSerThrAlaProArgArgSerLysLysLeuAspAlaSerGln 892  
QY 991 GCTCAGGAC-----ATCAAGATGAGGAGCCTGGAGACTTGGGCGCAGCGAGCGT 1041  
Db 893 AsnAsnAspProAspAlaSerGluSerGlnGluAspAsp----- 906  
QY 1042 GAATGTGAGGTACGACCCCAAGCCCTGTATTCATTTGCCGTCAGGCTCAGACAC 1101  
Db 907 -----AspProAsnLysLeuTrpCysIleCysArgGlnProHisAsnAsn 921  
QY 1102 AGGTTTATGATTGCTGACCGCTGTGAAGATCGTTTCATGGCGATTTGTGGGCATT 1161  
Db 922 ArgPheMetIleCysCysAspLeuCysGluAspTrpPheHisGlyThrCysValGlyVal 941  
QY 1162 TCTGAGGCTCGAGGAGGCTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCAACTGC 1221  
Db 942 ThrLysAlaMetGlyThrAspMetGluAsnLysGlyIleAspTrpLysCysProLysCys 961  
QY 1222 ACCATTCTCAAGTCAGGATGAGACTCATTCAAGAACGGCAGATCAGCAGCAAGCTAAA 1281  
Db 962 -----ValLysArgGlnGluGluArg 968  
QY 1282 TGGAGACCTGGAGATGCTGATGGCCCGATTTGTACAGTATATAGGAACAATAGAGCAGAG 1341  
Db 969 SerGlnProArgIleThrAspMetLeuValThrArgProThrThrGlnProGluGlnArg 988  
QY 1342 TCTACGAGAGCAGCAGGATA-----AAGGTAGATTTGAGAAAGCTCCAAATCCAAGT 1395  
Db 989 ProSerGluThrLysValLeuThrThrAlaGluIleValGlnValAlaAlaProSer 1008  
QY 1396 GGC---AAGAAGAACTCAAGATCTCCAGGCTGTGATAGAGGCGCCT----- 1440  
Db 1009 AlaProArgArgThrLeuProValValLeuThrValAlaSerSerProMetArgIlePro 1028  
QY 1441 -----GGTGCCTCAAAA----- 1452  
Db 1029 MetAlaLysProAlaLysPheProThrGlyAlaIleSerHisGlnGlnGlnGln 1048

1453 -----TGATTGGCCCGGG----- 1467  
QY  
1049 LeuAsnPheIleArgLeuGlyProSerProGlyLysArgIleSerGluThrLeuCysVal 1068  
Db  
1468 -----TGCTGTACGTGCGCGCCGCGACTCGGTCTACTGAGTAATGACTGTATCTCTCAA 1524  
QY  
1069 ValCysLysArgProAlaSerThrSerSerValTyrCysGlyGluCysIleArgLys 1088  
Db  
1525 CACGCC-----GCA 1533  
QY  
1089 TyrAlaGlnSerAlaIleGlnAlaHisAlaAlaThrLysGlyProLeuProGlnAsnAla 1108  
Db  
1534 GCGACAAATGAAGTTTCTAGCTCAGTAAAGACACAGCCAAAGCTTAAAGAAAGATG 1593  
QY  
1109 GlyAlaGlnSerLeuLeuAsnAsnSerPheAspAlaLysLysAsnLysLysAspLeu 1128  
Db  
1593 ----- 1593  
QY  
1129 PheGluAspValLeuArgGlnAlaAspThrValSerLysValGluArgIleAsnValPhe 1148  
Db  
1593 ----- 1593  
QY  
1149 GluArgLysSerGlyArgValIleThrGlyHisMetAlaProSerAlaHisGlnPheArg 1168  
Db  
1594 AAGATGAGCCAGAGAGCCAGT-----CTTCCGAAATGCGGTCTCAGCGCAGGT 1644  
QY  
1169 LysTrpLeuGlnGluAsnProSerPheGluValLeuProSerGlyThrValGlnSer--- 1187  
Db  
1645 ATTTAAATCTTCTGTGTCACAGAGA-----CCAGCTCAGAAAAAAGAGACC 1695  
QY  
1188 -----AlaAspAlaGluLysArgLeuLysGlyAlaProGluAlaAlaThrSer 1204  
Db  
1696 ACATGAGAGGAGGAGTGTCTCCGCGGAGT-----GAAGCACTCGGAGAGAA 1749  
QY  
1205 ThrSerGluProAlaValLeuGlyValAlaLysLysProGluGlyProAlaLysLeu 1224  
Db  
1750 GCAGCTGTGAGACAGCAGCGCTCGTGGCGGAGGATCAATCAATGAGTAAAG 1809  
QY  
1225 SerHisProGlnAsnThrThrValGlnAlaSerHisGlnLeuGlyIleSerSerValArg 1244  
Db  
1810 CCA-----GAAACACTGCT-----GCTCCTCGCGCTCA 1839  
QY  
1245 ProLeuAlaLysLysAspLysGluLysThrProThrValGlnAlaProThrProAsn 1264  
Db  
1840 CTGTTGTATAATGTATGATCACTAGGCGTGGCTCTCGAGACCCCTCCCGT 1893  
QY  
1265 -----ArgIleAlaLysLysProGluProValArg 1275  
Db  
Q9W352 PRELIMINARY; PRT; 663 AA.  
AC Q9W352;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE CG17446 protein.  
DE CG17446  
GN CG17446  
OS Drosophila melanogaster (fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,  
RA Wan X.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Bottier P., Bottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A.,  
RA Cherry J.M., Cawley S., Dahke C., Davenport L.B., Davies P.,  
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Periera S., Fleischmann W.,  
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Hostin N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibgwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,  
RA Swirekas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.W., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RA "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
DR EMBL; AB003446; AAF46483.1; -.  
DR FlyBase; FBgn0030121; CGI17446.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
DR InterPro; IPR002857; Znf\_CXAC.  
DR InterPro; IPR001985; Znf\_PHD.  
DR Pfam; PF00628; PHD; 1.  
DR Pfam; PF02008; zf-CXAC; 1.  
DR SMART; SM00249; PHD; 1.  
DR PROSITE; PS01359; ZF\_PHD\_1; 1.  
DR PROSITE; PS50016; ZF\_PHD\_2; 1.  
SQ SEQUENCE 663 AA; 76695 MW; 8573CD209AE28157 CRC64;  
  
Alignment Scores:  
Pred. No.: 1,19e-06 Length: 663  
Score: 217.00 Matches: 88  
Percent Similarity: 32.56% Conservative: 53  
Best Local Similarity: 20.32% Mismatches: 122  
Query Match: 4.63% Indels: 170  
DB: 5 Gaps: 15  
  
US-09-787-016A-1 (1-2610) x Q9W352 (1-663)  
QY 907 AGTAAGCAGGAGCCCGAGACGATCAGGGGTGTGTCCAGGCTGGGAAGATGACAGA 965  
Db 28 SerLysGlnLysGluGluIleArgArgGluAlaArgLysPheAspLeuProGluArg 47  
QY 967 GAGAGTAAGTTGAGGAGAAAGCGGCTCAGACATCAAGATGAGGAGCTTG 1026  
Db 48 LysSerLysIle-----AlaThrIleLeuLysGlnGluAspGlnAla----- 61  
QY 1027 GCGCGACCGAGGCTGATGAGGTTACGACCCCAACCGCTGTATTCATTGCGC 1086  
Db 62 -----TyrCysIleCysArg 66  
QY 1087 CAGCCTCACACAAACAGGTTTATGATTGCTGACCGCTGTGAGAGATGTTTCATGCG 1146  
Db 67 SerSerAspCysSerArgPheMetIleGlyCysAspGlyCysGluGluTyrTyrHisGly 86  
QY 1147 GATTGTGTGGGCTTCTTGTAGGCTCAGGAGGCTTTTGGAAAGGATGGGAGAGCTAT 1206  
Db ----- 1147

Db 87 AspCysIleGlyThrGluLysGluAlaLysHisIleLysGln-----Tyr 102  
 QY 1207 ATCTGCCCAAACTGC-----ACCATCTGCAG----- 1233  
 Db 103 TyrCysArgCysLysLysGluAsnProGluLeuGlnThrIlePheArgLeuValAla 122  
 QY 1233 ----- 1233  
 Db 123 ThrGluArgAlaAlaAlaSerAsnAlaAlaSerThrSerLeuAsnAlaProGlyValGly 142  
 QY 1234 -----GTGCAGGATGAGACTCATTTCAGAAACGGCA 1263  
 Db 143 ProSerGlyAlaAlaProAlaAlaAlaProValAlaSerAlaThrThrSerGlnGlnAla 162  
 QY 1264 -----GATCAGCAGGAACTAAATGG 1284  
 Db 163 ProProProThrThrAlaAlaAlaLysArgLysAsnSerSerAlaArgGluProLysMet 182  
 QY 1285 -----AGACCTGGAGATCCTGATGC----- 1305  
 Db 183 GlyLysArgCysGlyThrCysGluGlyCysArgArgProAsnCysAsnGlnCysAspAla 202  
 QY 1306 -----ACCATTTGTTCAAGTATA 1323  
 Db 203 CysArgValArgValGlyHisLysProArgCysIlePheArgThrCysValValGlnAla 222  
 QY 1324 GGAACATAGACGAGAGTCTAGCGAGACCAA---GGATTAAGCGTAGATTTCAGAAA 1380  
 Db 223 AlaThrValLeuLysGluSerGlnAlaThrGlnAlaGlyProSerArgLysArgGluLys 242  
 QY 1381 GTGTCAATCCAGTGGCAGAGAAATCTC-----AAGATCTCCAGCTGTGATA--- 1431  
 Db 243 AlaAlaProLysSerArgAsnValGlnValGlyProArgAlaAlaSerProGluIlePhe 262  
 QY 1432 -----GAGCGCTGCTGCTCAAAATGATTGGCCCGTGTGCTCACTGCGG 1482  
 Db 263 LeuAsnProGluGlnGlyIleArgGlnCysTyrGlyProAsnCysCysSerHisAla 282  
 QY 1483 CAGCCCGATCTCGTGTACTGCACTGACTGT----- 1515  
 Db 283 ArgProGlnSerLysTyrCysSerAspLysCysGlyPheAsnLeuAlaThrLysArgIle 302  
 QY 1515 ----- 1515  
 Db 303 PheGlnValLeuProGlnArgLeuGlnGluTrpAsnLeuThrProSerArgAlaAlaGlu 322  
 QY 1516 -----ATCTCAACACGCGCGACGACATGAAGTTTCTAAGC 1554  
 Db 323 GluThrArgLysHisLeuAspAsnIleArgHisLysGlnSerLeuValArgPheAlaLeu 342  
 QY 1555 TCAGGTAAGACAGAACGCAAGCTTAAGAAAGATGAAGATGACGACGACGAGAGCC 1614  
 Db 343 AlaGluLeuGluLysArgSerGluGluLeuAsnMetValValGluArgAlaLysArgSer 362  
 QY 1615 AGTCTCCGAATCGGTGCTCAGCGAGGTATTAATCTCT----- 1656  
 Db 363 SerIleAspThrLeuGlySerGlnAspThrAlaAspMetGluAspGluGlnSerMetTyr 382  
 QY 1657 -----TCTGTGCACAGACGACGCTCCAGAA----- 1683  
 Db 383 CysIleThrCysGlyHisGluIleHisSerArgThrAlaIleLysHisMetGluLysCys 402  
 QY 1684 -----AAAAAGACACACAGTACGAGAGCGAGTGTGTCTCGCGGAGTACGAGCA 1737  
 Db 403 PheAsnLysTyrGluSerGlnAlaSerPheGlySerIlePheLysThrArgMetGlu--- 421  
 QY 1738 CTCGGTAAGACGACGCTGTGACAGCAGCAGCGCGCTCG 1776  
 Db 422 ---GlyAsnAsnMetPheCysAspPheTyrAsnProAla 433

RESULT 3

Q9VUB5

ID Q9VUB5 PRELIMINARY; PRT; 3146 AA.

AC Q9VUB5;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE CG9007 protein.  
 GN CG9007  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 PC STRAIN=Berkeley.  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazei R.G., Champe C.R., Miklos G.L.G.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Baldwin D.,  
 RA Abril J.P., Agbayani A., An H.-J., Andrews-Pfannkoch C., Basley E.M.,  
 RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brattier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K., Ewing-Lewis C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fodor C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,  
 RA Glöckle A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,  
 RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Liu X., Mattei B., Levisky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Wang X.,  
 RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Banon J., An H., Baldwin D., Banon J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
 RA Ferreira S., Frise B., Galle R.F., Garg N.S., George R.A.,  
 RA Gonzales C., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibegwan C., Jallali M., Kuse D., Li P., Mattei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B., Scheeler F.,  
 RA Phuanavong S., Pittman G.S., Puri V., Richards R., Tector C., Tyler D.,  
 RA Stapleton M., Strong R., Svirskaas R., Tector C., Tylor D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RT "Sequencing of Drosophila melanogaster genome."  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,



RA Hradecky P., Huang Y., Kaminker J.S., Prochownik S.B., Smith C.D.,  
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,  
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,  
 RA Searle S.M.J., Smith B., Shu S., Smutniak P., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.B.;  
 RT Annotation of *Drosophila melanogaster* genome.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RN SEQUENCE FROM N.A.  
 RP Flybase;  
 RA Flybase;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: CONTAINS 1 SET DONAIN.  
 DR EMBL; AB003536; AAF49773.3; -;  
 DR FlyBase; FBgn0036398; CG9007.  
 DR GO; GO:0003677; F.DNA binding; IEA.  
 DR GO; GO:0006355; P.regulation of transcription, DNA-dependent; IEA.  
 DR InterPro; IPR001214; Znf.  
 DR InterPro; IPR001965; Znf\_PHD.  
 DR Pfam; PF00628; PHD; 1.  
 DR Pfam; PF00856; SET; 1.  
 DR SMART; SM00317; SET; 1.  
 DR SMART; SM00249; PHD; 1.  
 DR PROSITE; PS00280; SET; 1.  
 DR PROSITE; PS01359; ZF\_PHD\_1; 1.  
 DR PROSITE; PS00166; ZF\_PHD\_2; 1.  
 SQ SEQUENCE 3146 AA; 336443 MW; 81EEAC5C308PEC31 CRC64;

## Alignment Scores:

Pred. No.: 3,72e-06 Length: 3146  
 Score: 212.00 Matches: 154  
 Percent Similarity: 31.92% Conservative: 79  
 Best Local Similarity: 21.10% Mismatches: 254  
 Query Match: 4.52% Indels: 243  
 DB: 5 Gaps: 33

US-09-787-016a-1 (1-2610) x Q9VBUS (1-3146)

QY 4 GTTGCGCGTCCGCGCCACTCCGGG-----CGTTGGGGAAATGG 42  
 DB 487 GlyAsnMetProProProThrGlnIleTyrSerAsnSerThrSerAsnGlyAlaAla 506  
 QY 43 CTGCGAGACCTAGAGCGCTGGGAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGA 102  
 DB 507 AlaSerSerProGlyGlyAsnAlaSer-----GlyAsnMetLeuLeuAlaHisTyr 523  
 QY 103 GTTGTGAAATACGAGCCTGTACTCTGTGACAGTGGCTGACA----- 147  
 DB 524 GlnAlaAlaGlyThrLysProValSerSerAlaSerPheIleThrValThrGlyThrPro 543  
 QY 148 -----ACAGTGTGTGTGAGCCTGGCTGTCTGTGGACCCAGAGGTTTCTGTGCCA 201  
 DB 544 ProValThrVal-AlaThrThrProSerValSerIleSerHisGlyPheAla----- 561  
 QY 202 GGGTTTTCGTGTATTAGGATTTTCAGGGAAGAGTGTCAAGCTTTTCAAGTGTGGAGCA 261  
 DB 562 -----SerGlySerAlaAlaIleSerSer----- 569  
 QY 262 GGTATGGACGACAAAGCGACCGGACCAATGAGGAGGACCTAAGCCATCAACCCACC 321  
 DB 570 ----TyrMetSerSerAlaThrAlaAlaArgArgGlnSerValSer----- 583  
 QY 322 AGCAAGAGTTCAGGAACATGGGGTTTTCGAAGGAGCCACTATCGCCAAAGAGGAGGC 381  
 DB 584 -----AlaProSerSerArgAl 589  
 QY 382 GCAGGGGACCGGAGGCTGACCCACTGAGCGCCGCCACCCACAGCAGCAGCTGGCGCTG 441  
 DB 589 aValSerLeuGluArgLysGlnHisGlnGlnLeuGlnHisAsp----- 604

QY 442 TCCCTGCGCGCAGTGGGAGGC-----AGCCAAAGCGCACTGAGGCGCTGGAGCAG 492  
 DB 605 -----ValIleGlyGlyArgLysAlaProThrValIleGluTyrTyrAsnLys 621  
 QY 493 T----- 493  
 DB 621 SHisGlyValAsnSerIleValGlySerSerAsnAsnLeuAlaGlnSerAsnSerMetSe 641  
 QY 494 -----TCTGACCATTCGCGCGCGCGCGCGAGGAGCATGC----- 532  
 DB 641 rAsnLeuAlaGlyProArgSerAsnSerGlySerGlyPheAlaThrThrThrProThrPr 661  
 QY 533 -----CTGCTCCCTGGAGGATTCCTGGTGAGCCCGCCACGCTCTCTGCCCGCC 576  
 DB 661 cAlaThrProLeuHisLeuThrProValAsnValProValHisValGluAlaAlaProPr 681  
 QY 577 ACAGACGCGC-----AGACAGCCTCCGAGGGCAGCGCTGGAAACGCGCT 618  
 DB 681 cSerSerProAlaLeuValLysGlySerSerGlnProPro-----AlaGlnProGlnGl 699  
 QY 619 TCTGAGACCAAGACCGCGCCCGCTCTCTCCACAGCTGTGAAGGAAGACCCAGCCTCT 678  
 DB 699 nGlnGlnGlnGlnAlaHisProLeuGlyProAsnGlnLeuAsnAlaAsnAspGluGlu 719  
 QY 679 TCTGAAAAGGTGAAAGGAGGATGACCACCATGACACT---CCGATAGTCACAGCGAT 735  
 DB 719 uTyr-----IleGluGluValArgProValProValLeuThr--Gln 732  
 QY 736 GCGCTGACCTTGAAGAGCTTCAGAAATCGCCTTCGC-----AGGAAG 777  
 DB 733 AspLeuArgLeuGlnGlnLeuHisAlaIleMetGlnAspHisThrTyrAlaSerGlnGln 752  
 QY 778 CGGGAACAGGAGCGC-----ACTGAGAGCGCCCTGAAAGGGATCCAG 819  
 DB 753 GlnGlnGlnGlnProGlnGlnAlaAlaGlyAspThrThrAsnProGlyAlaAlaGlnGln 772  
 QY 820 AGTCGCTGCGGAAGAACCGCGGAGGAGGTCCCGCGAGACTGTG---GGCTCCGAG 876  
 DB 773 ValGlnGlnProGlnGlnTrpSerLeuGlyGlyIleGlyValThrValSerGlySerGln 792  
 QY 877 GCCAGTGACATGTGGAGGCGTCTGCCAGTAAGCAGGAGCCCGAGACGATCAGGGG 936  
 DB 793 GlyThrProThrAlaValGlyGlyTyrCysSer-----TyrPheGlyGlnGln 808  
 QY 937 GTTGTGTCCAGGCTGGGAAGATACACAGAGAGTAAATGTTGAG----- 981  
 DB 809 IleAlaArgSerGlnAlaAspAspAlaHisSerAlaIleSerSerSerSerArgMet 828  
 QY 982 CGAAAGCGGCTCAGGACATCAAGATGAGGAGCCTCGAGACTTGGGCGGACCGAGCCT 1041  
 DB 829 GlyLeuAlaSerThrAspIle-----AspProGlyGluGluThrGluThrAlaPro 845  
 QY 1042 GAATGTGAGGTTTACGACCCCAACGCCCTGTATTGCAATTCGCGCACCTCACAACAAC 1101  
 DB 846 GluAlaGluAlaGluAspSerValThrArgCysIleCysGluLeuThrHisAspAsp 865  
 QY 1102 AGGTTTATGATTTCTGTACCGCTGTGAAGATGTGTTTCATGCGCATGTGTGGCGATT 1161  
 DB 866 GlyTyrMetIleCysCysAspLysCysSerAlaTrpGlnHisValAspCysMetGlyIle 885  
 QY 1162 TCTGAGGCTCGAGGAGGCTTTTGGMAAGGAATGGGGAAGACTATATCTGCCAACTGC 1221  
 DB 886 AspArg-----GlnAsnIleProGluGluTyrMetCysGluLeuCys 899  
 QY 1222 ACCATTCTGCAAGTGCAGATGAGACTCATTCAGAAACGGCAGATCAGCAGGAAGCTAAA 1281  
 DB 900 GlnProArgAlaVal---AspLysAlaArgAlaArgAlaLeuGlnArgGlnLysArgLys 918  
 QY 1282 TGGAGA-----CCTGGAGATGCTGATGCCACCGATTGTACAGTATAGA 1326  
 DB 919 GluHisMetLeuLeuValAlaThrGlnAlaAlaAsnGlyAlaAlaValAlaAlaGly 938



DR ENBL; AE003446; AAP46482.2; -  
DR FlyBase; FBgn0030120; CGI17440.  
DR GO; GO:0003677; P:DNA binding; IEA.  
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
DR InterPro; IPR001965; Znf\_PHD.  
DR Pfam; PF00628; PHD; 1.  
DR SMART; SM00249; PHD; 1.  
DR PROSITE; PS01359; ZF\_PHD\_1; 1.  
DR PROSITE; PS00046; ZF\_PHD\_2; 1.  
DR SSSEQUENCE 366 AA; 42048 MW; CB93AAFEAC514CF3 CRC64;

Alignment Scores:  
Pred. No.: 1,07e-05 Length: 366  
Score: 202.50 Matches: 63  
Percent Similarity: 37.55% Conservative: 29  
Best Local Similarity: 25.71% Mismatch: 82  
Query Match: 4.32% Indels: 71  
DB: 5 Gaps: 8

US-09-787-016A-1 (1-2610) x Q8T3Y1 (1-366)

QY 910 AAGCAGAGCCCGAGACGATCAGGGGGTGTGTCCAGCGCTGGAAAGATGACAGAGAG 969  
DB 7 LysThrGluGluLeuArgArgGluLeuAlaArgGluPheAspLeuProGluArgLys 26  
QY 970 AGTAAGTTGGAGGGAAGAGCGGCTCAGGACATCAAGATGAGGAGCTGGAGACTGGGC 1029  
DB 27 SerLysLeuAlaThrLeuLysGln-----LysAspArgGlu----- 39  
QY 1030 CGACCGAAGCCTGAATGTGAGGGTTACGACCCCAAGCCCTGTATTGCATTTGCCGCCAG 1089  
DB 40 -----TyrCysIleCysArgSer 45  
QY 1090 CCTCACAAACACAGGTTTATGATTGCTGTGACGGCTGTGAAGAAAGTGTTCATGCGCAT 1149  
DB 46 SerAspCysSerArgPheMetIleGlyCysAspGlyCysGluGluTrpTyrHisGlyAsp 65  
QY 1150 TGTGTGGCATTTCTCAGGCTCGAGGGAGCGCTTTTGGAAAGGAATCGGGAAGACTATATC 1209  
DB 66 CysIleGluLeuThrGluLysAspAlaGluHisLe-----LysAsnTyrTyr 81  
QY 1210 TGGCCCAAACTGC-----ACCATTCGCAAGTGCAGGATGAG 1245  
DB 82 CysArgArgCysIlyLysLysGluAsnProGluLeuGlnThrIlePheArgLeu----- 98  
QY 1246 ACTCATTCACAAACGGCAGATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGCG 1305  
DB 99 ValAlaThrGluArgAlaAlaAlaSerAsnAlaAlaSerThrSerLeuAsnAlaProGly 118  
QY 1306 ACCGATTGTACAAATATAGGAACATA----- 1332  
DB 119 ValGlyProSerGlyAlaAlaProAlaAlaAlaProValAlaProAlaThrThrSerGln 138  
QY 1333 -----GAGCAG 1338  
DB 139 GlnAlaProProThrThrAlaAlaAlaLysArgLysAsnSerSerAlaGlnGluPro 158  
QY 1339 AAGTCTAGCCAGAACCAAGCGATAAAGGGTAGAATTACAGAAAGCTGCAAATCCAACTGGC 1398  
DB 159 LysGluSerGlnProThrGlnAlaGlyThrLysArgAspLysAlaAlaProLysThrSer 178  
QY 1399 AAGAGNAATC-----AAGATCTCCACCTGTGATA-----GAGCGCGCT 1440  
DB 179 AsnValGlnValSerProArgAlaValSerProGluIlePheLeuAsnProGluLeuGln 198  
QY 1441 GGTGCCTCAAAATGATTTGGCCCGGGTGTGTTCAGTGGCGAGCGCCGACTCGGTGTAC 1500  
DB 199 GlyIleGlnGlnCysHisGlyProAsnCysCysSerHisAlaArgProGlnSerLysTyr 218  
QY 1501 TGCAGTAATGACTGT 1515  
DB 219 CysSerAspGluCys 223

RESULT 5

Q9W0T2	PRELIMINARY	PRT; 2649 AA.
ID	Q9W0T2	
AC	Q9W0T2	
DC	01-WAY-2000 (TrEMBLrel. 13, Created)	
DT	01-OCT-2002 (TrEMBLrel. 22, Last sequence update)	
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)	
DE	CG32346 protein.	
GN	E(BX) OR CG7022 OR CG10894 OR CG17135 OR CG32346 OR CG32478.	
OS	Drosophila melanogaster (Fruit fly).	
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Peerygota;	
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	
OC	Ephydroidea; Drosophilidae; Drosophila.	
OX	NCBI_TaxID=7227;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RP	STRAIN=Berkley;	
RX	MEDLINE=20196006; PubMed=10731132;	
RA	Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,	
RA	Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,	
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,	
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,	
RA	Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,	
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,	
RA	Ball J.F., Abghyani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,	
RA	Bralley R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,	
RA	Beeson K.Y., Benoe P.V., Berman B.P., Bhandari D., Bolshakov S.,	
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,	
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,	
RA	Cherry J.M., Cawley S., Dallic C., Davenport L.B., Davies P.,	
RA	de Pablo E., Deicher A., Deng Z., Mays A.D., Dew I.J., Dietz S.M.,	
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,	
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,	
RA	Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,	
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,	
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,	
RA	Hoskin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,	
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,	
RA	Kimmel B.E., Kodira C., Kraft C., Kravitz S., Kulp D., Lai Z.,	
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,	
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,	
RA	Merkulov G., Milhina N.V., Mobarry C., Morris J., Moshrefi A.,	
RA	Mout S.M., Moy M., Murphy B., Murphy L., Muzay D.M., Nelson D.L.,	
RA	Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacleb J.M.,	
RA	Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,	
RA	Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,	
RA	Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,	
RA	Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,	
RA	Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,	
RA	Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,	
RA	Williams S.M., Woodgate T., Worley K.C., Wu D., Yang S., Yao Q.A.,	
RA	Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,	
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu S.,	
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,	
RT	"The genome sequence of Drosophila melanogaster.";	
RL	Science 287:2185-2195(2000).	
RL	[2]	
RP	SEQUENCE FROM N.A.	
RA	Celiker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,	
RA	Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,	
RA	Branson J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,	
RA	Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,	
RA	Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,	
RA	Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,	
RA	Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,	
RA	Ibegwan C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,	
RA	McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,	
RA	Pacleb J., Paragov V., Park S., Patel S., Pfeiffer B.,	
RA	Phouanavong S., Pittman G.S., Puri V., Richards S., Schesler F.,	
RA	Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,	
RA	Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.,	
RT	"Sequencing of Drosophila melanogaster genome.";	
RL	Submitted (MAR-2000) to the ENBL/GenBank/DBJ databases	

[3]  
SEQUENCE FROM N.A.  
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
RA Tupy J.H., Bergman C., Carlson J.W., Celnik S.E.,  
RA Clump M., Drysdale R., Emmert D., Friese E., de Grey A., Harris N.,  
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
RA Seale S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
RA "Annotation of Drosophila melanogaster genome.";  
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
[4]  
SEQUENCE FROM N.A.  
RA Adams M.D., Celnik S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
[5]  
SEQUENCE FROM N.A.  
RA FlyBase;  
RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
RL EMBL; AF003467; AAF47361.2; -;  
DR FlyBase; FBgn000541; E(hx).  
DR GO; GO:0005634; C:nucleus; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:000489; P:electron transporter activity; IEA.  
DR GO; GO:0006118; P:electron transport; IEA.  
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
DR InterPro; IPR006337; AT hook.  
DR InterPro; IPR001487; Bromodomain.  
DR InterPro; IPR000345; CytC heme\_BS.  
DR InterPro; IPR004022; DDT\_dom.  
DR InterPro; IPR006209; EGF-like.  
DR InterPro; IPR001965; Znf\_PHD.  
DR Pfam; PF02178; AT hook; 1.  
DR Pfam; PF00439; bromodomain; 1.  
DR Pfam; PF02791; DDT; 1.  
DR Pfam; PF00628; PHD; 3.  
DR PRINTS; PRO0503; BROMODOMAIN.  
DR SMART; SM00297; BROMO; 1.  
DR SMART; SM00249; PHD; 3.  
DR PROSITE; PS00633; BROMODOMAIN\_1; 1.  
DR PROSITE; PS00114; BROMODOMAIN\_2; 1.  
DR PROSITE; PS00190; CYTOCHROME\_C; 1.  
DR PROSITE; PS01186; EGF\_2; 2.  
DR PROSITE; PS01359; ZF\_PHD\_1; 1.  
DR PROSITE; PS00016; ZF\_PHD\_2; 2.  
SQ SEQUENCE 2649 AA; 298507 MW; CODF7E015EA5403 CRC64;

Alignment Scores:  
Pred. No.: 3.13e-05 Length: 2649  
Score: 198.50 Matches: 87  
Percent Similarity: 36.75% Conservative: 78  
Best Local Similarity: 19.38% Mismatches: 147  
Query Match: 4.23% Indels: 139  
DB: 5 Gaps: 16

US-09-787-016A-1 (1-2610) x Q9W0T2 (1-2649)  
QY 362 CTATGCCAAGCGAGGGCGGAGCGGAGCGTACCCACCTGGAGCGCCACCC- 420  
Db 2180 IletThrAsnGluSerPheAlaGlyThrSerSerLeuLeuGluGlySerGluHisAsp 2199  
QY 421 ---CCACAGCAGCAGCTGGCGCTCTCCCTCGCGCGAGTGGGAGGAGCCCAAGCGCACT 477  
Db 2200 GluProThrAsnLeuAlaGlyLeuAspIleSerGluThrAspLeuGluAsnLysGlnAsn 2219  
QY 478 GAGGCG-----GTGAGCAGTCTCTGACCATTCGACCATTCGCGCGCGCG 516  
Db 2220 GluSerPheValThrArgGlyTyrIleGlnLysSerIleSerAsnAlaLeuLysGln 2239  
QY 517 GGAGGAGGAGCGAGCGCTGTCTCCCTGGAGGAT-----TCTGGTGGAG 558  
Db 2240 Gly-----AsnLeuSerProGluLeuGluGluLysLeuValCysMetGlnLysGlnGln 2257

QY 559 CCCACGTCCTGCCCGCCGACAGACGCGGAGCGAGCCTCCGAGGCGCAGCGTGAAGAGCGCT 618  
Db 2258 GluAsnAlaAsnSerThrAsnGluThrGlyThrCysSerArgGlySerValAsn----- 2275  
QY 619 TCTGAGACGAGACGCGCCCGCTCTCCACAGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAG 678  
Db 2276 -----GluGluAlaLeuThrProSerArgGlnThrAspAspThr 2288  
QY 679 TCTGAAAGAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 738  
Db 2289 GluTrpLysIleArg-----ThrSerLeuArgArgProAsnAla 2301  
QY 739 CTGACCTTGAAGAGCTTTCAGATCGCTTCGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 798  
Db 2302 MetThrThrSerSerGlnPheAsnArgIleLeuLysLysAsnArgSerLysAsnAspGlu 2321  
QY 799 AGGCGCTGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 858  
Db 2322 ValAlaGluLeuGlyGluGlnLysGlnSerGlnLeuGluArgHisLysGluLeuLys 2341  
QY 859 GAGACTGTG-----GGCTCGGAGCGCCAGTGC 885  
Db 2342 LysAsnIleLeuArgLysArgSerLeuLeuGluArgAsnLeuGlnSerGluIleHisGlu 2361  
QY 886 ACTGTGAGGCGCTCTCCCGAGTAAGCAGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 945  
Db 2362 AspValLysThrLysValGlnArgHisValArgPro-----LeuSer 2375  
QY 946 CAGGCTGGGAAAGATGACAGA---GAGAGTAAGTTGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1002  
Db 2376 AsnAlaSerProAspGluGlnSerGluAsnGluArgSerGlyGluProAsnLeuAspPhe 2395  
QY 1003 AAAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1041  
Db 2396 LysArgThrGluValGlnAsnProArgHisGlyAlaGlyArgProLysLysLeuThrArg 2415  
QY 1041 ----- 1041  
Db 2416 LysLysGluLysLeuTyrCysIleCysArgThrProTyrAspAspThrLysPheTyrVal 2435  
QY 1041 ----- 1041  
Db 2436 GlyCysAspLeuCysSerAsnTrpPheHisGlyAspCysValSerIleThrGluGluAla 2455  
QY 1042 -----GAAATGTGAG---GGTTACGACCCCAACGCC 1068  
Db 2456 SerLysLysLeuSerGluPheIleCysIleAspCysLysArgAlaArgGluThrGlnGln 2475  
QY 1069 CTGTATTCATTTGCCCGACGCTCACAAAC---AACAGTTTATGATTGCTGTCACCGC 1125  
Db 2476 LeuTyrCysSerCysArgGlnProTyrAspGluSerGlnPheTyrIleCysCysAspLys 2495  
QY 1126 TCTGAAGAATGCTTTCATGCGGATTTCTGGGCTTCTGAGGCTCGAGGAGGAGGAGGAGGAG 1185  
Db 2496 CysGlnAspTrpPheHisGlyArgCysValGlyIleLeuGlnSerGluAlaGluPheIle 2515  
QY 1186 GAAAGGAAATGGGAGAGACTATATCTGCCAACTGCACCATTTCTCAAGTGCAGGATGAG 1245  
Db 2516 -----AspGluTyrValCysProGluCysGln----- 2524  
QY 1246 ACTGATTGAGAAACGCGAGATCAGCAGGAGCTAATGAGACCTGGAGACTCTGATGCG 1305  
Db 2525 -----ArgLysAsnAspAlaAsnAla 2531  
QY 1306 ACCGATTGTACAAGTATAGGAACAATAGACAGAGAGTCTAGCGAGACCAAGCGGATAAAG 1365  
Db 2532 AlaAsnMetLysLysLeuThrSerAsnAspValGluGluLeuLysAsnLeuIleLysGln 2551  
QY 1366 GTAGAAATTTGAGAAAGCTGCA-----AATCCAGTGGCAGAGGAGAACTCAAG 1413  
Db 2552 MetGlnLeuHisLysSerAlaTrpProPheMetGluProValAspProLysGluAlaPro 2571  
QY 1414 ATCTTCCAGCTGTGATAGAGGCGCT 1440

Db 2572 AspTyrTrpLysValLeuLysGluPro 2580  
 Q95VB8 PRELIMINARY; PRT; 2669 AA.  
 AC Q95VB8;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Nucleosome remodeling factor large subunit NURF301.  
 GN E(BX) OR NURF301 OR CG7022 OR CG10894 OR CG17135 OR CG32346 OR CG32478.  
 OS Drosophila melanogaster (fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=2146388; PubMed=11583616;  
 RA Xiao H., Sandatzopoulos R., Wang H., Hamiche A., Ranallo R., Lee K.,  
 RA Fu D., Wu C.;  
 RT "Dual functions of largest nurf subunit nurf301 in nucleosome sliding  
 RT and transcription factor interactions";  
 RL Mol. Cell 8:531-543 (2001).  
 DR EMBL: AF417921; AAL16644.1; --  
 DR FlyBase: FBgn000541; E(bx).  
 DR GO: GO:0005634; C:nucleus; IEA.  
 DR GO: GO:0003677; F:DNA binding; IEA.  
 DR GO: GO:0005489; F:electron transport activity; IEA.  
 DR GO: GO:0006118; P:electron transport; IEA.  
 DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
 DR InterPro: IPR000637; AT hook.  
 DR InterPro: IPR001487; Bromodomain.  
 DR InterPro: IPR000345; CycC\_heme\_BS.  
 DR InterPro: IPR004022; DDT\_dom.  
 DR InterPro: IPR006209; EGF-like.  
 DR InterPro: IPR001965; Znf\_PHD.  
 DR Pfam: PF02178; AT hook; 1.  
 DR Pfam: PF00439; bromodomain; 1.  
 DR Pfam: PF02791; DDT; 1.  
 DR Pfam: PF00628; PHD; 3.  
 DR PRINTS: PR00503; BROMODOMAIN.  
 DR SMART: SM00297; BROMO; 1.  
 DR SMART: SM00571; DDT; 1.  
 DR SMART: SM00249; PHD; 3.  
 DR PROSITE: PS00633; BROMODOMAIN 1; 1.  
 DR PROSITE: PS00014; BROMODOMAIN 2; 1.  
 DR PROSITE: PS00180; CYTOCHROME\_C; 1.  
 DR PROSITE: PS01186; EGF\_2; 2.  
 DR PROSITE: PS01359; ZF\_PHD\_1; 1.  
 DR PROSITE: PS00016; ZF\_PHD\_2; 2.  
 DR PROSITE: PS00016; ZF\_PHD\_2; 2.  
 SQ SEQUENCE 2669 AA; 300687 MW; 654925AFFF489D6F CRC64;

Alignment Scores:  
 Pred. No.: 3.13e-05 Length: 2669  
 Score: 196.50 Matches: 87  
 Percent Similarity: 36.75% Conservative: 78  
 Best Local Similarity: 19.38% Mismatches: 147  
 Query Match: 4.23% Indels: 138  
 DB: Gaps: 16

US-09-787-016A-1 (1-2610) x Q95VB8 (1-2669)

QY 362 CTATCCCAAGCAGAGGGCGGAGCGGAGCGCTGACCCACTGGAGCGCGCACCC- 420  
 Db 2200 IleThrTrpAsnGluSerPheAlaGlyThrSerSerLeuLeuGluGlySerGluHisAsp 2219  
 QY 421 ---CCACAGCAGCAGCTGGCGCTGCTCCCTCGCGCAGTGGGAGGCGCCAGCGCACT 477  
 Db 2220 GluProThrAsnLeuAlaGlyLeuAspIleSerGluThrAspLeuGluAsnLysGlnAsn 2239







```

Db 2422 leuAspLysIleAspLysGluGluGlnAlaAlaLysLysArgLysArgGluGlu 2442
QY 854 CCGCCAGACATGCGGCTCGAGCCAGTGCACACTGTGCGAGCGCTCCGCCAGTAAGC 913
Db 2442 erValGluGlnLysArgSerLysGlnAlaAlaThrLysLysSerAlaLeuLeuPheLysH 2462
QY 914 AGAGCCCGAGAACATGATCAGGGGTTGTGTCCAGGCTGGAAAGATGACAGAGAGTA 973
Db 2462 isLysGluGlnLeuArgAlaGluIleLeuLysLysArgAlaLeuLeuAspLysAspLeuG 2482
QY 974 AGTTGCGAG-----GGAAGGGCGCTCAGACATCAAGATGAGGAGCTGGAGACTTG- 1026
Db 2482 InileGluValGlnGluGluLeuLysArgAspLeuLysIleLysLysGluLysAspLeuM 2502
QY 1027 -----GCGCGAC 1033
Db 2502 etGlnLeuAlaGlnAlaThrAlaValAlaAlaProCysProProValThrProAlaProp 2522
QY 1034 CGAAGCTGAATGAGGGTTACGACCC----- 1062
Db 2522 roAlaProAlaProProProProProProProProProProProProProProProPro 2542
QY 1062 ----- 1062
Db 2542 euLeuSerThrProThrLeuProAlaAlaSerGlnLysArgLysArgGluGluLysA 2562
QY 1062 ----- 1062
Db 2562 spSerSerSerLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 2582
QY 1063 -----AACGCCCTGATTGCTGCGGAGCTCAGAAC-----AACAGGTTTATGATT 1114
Db 2582 ysAspThrLysLeuLysCysLysLysLysLysLysLysLysLysLysLysLysLysLys 2602
QY 1115 GCTGTACCGCTGTGAAGATGTTTCATCGCGATGTTGCGCATTTCTGCGCTCGAG 1174
Db 2602 lyCysAspArgCysGlnAsnTrpTrpHisGlyArgCysValGlyIleLeuGlnSerGluA 2622
QY 1175 GAGGCTTTTGAAGAGATGGGAGAGACTATATCTGCCCAACTGC 1221
Db 2622 laGluLeuIle-----AspLysLysValCysProGlnCys 2633

```

## RESULT 9

```

ID Q818F0 PRELIMINARY; PRT; 542 AA.
AC Q818F0;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DE Chromosome scaffold protein p85.
OS Euplotes crassus.
OC Eukaryota; Alveolata; Ciliophora; Spirotrichea; Hypotrichia;
OX Euplotida; Euplotidae; Moneuplotres.
NX NCBI_TaxID=5936;
RN [1]
RP SEQUENCE FROM N.A.
RA Sharp S.I., Pickrell J.K., Jahn C.L.;
RT "The identification of a novel chromosome scaffold protein that
RT associates with Tec elements undergoing en masse elimination in
RT Euplotes crassus."
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY15457; AAK7502.1; -.
SQ SEQUENCE 542 AA; 60264 MW; F0AB3B5E3AA771A1 CRC64;

```

## Alignment Scores:

```

Pred. No.: 0.000101 Length: 542
Score: 189.00 Matches: 113
Percent Similarity: 35.55% Conservative: 101
Best Local Similarity: 19.77% Mismatches: 204
Query Match: 4.03% Indels: 184
DB: 5 Gaps: 24

```

US-09-787-016A-1 (1-2610) x Q818F0 (1-542)

```

QY 211 GTTGTATTAGGATTTACGGGAAAAGTGTCCAAGCTTTCACTGTTGGAGCAGGTATGAC 270
Db 6 IleAlaThrLysValAlaGlyLysGlyLeuLysThr-----LysGlyLysLys 21
QY 271 GACAAAGCGCAGCCGCAATGAGGAGCCACTAAGGCCATCAACCCACC----- 321
Db 22 ThrLysAlaAlaGluLysSerLysLysIleGluLysAlaThrLysAspValGluMetLys 41
QY 322 -----AGCAAAAGAGTTTCAGAAAACATGCGGTTTTTCGAAGG 357
Db 42 AsnProSerAspGluGluAspIleAsnLysGluSerAlaArgThrSerHisThrAsnLys 61
QY 358 ACCACTATCCGACAGCAGAGCGCGAGGAGCGGAGCGGAGCTGACCCACTGGAGCGGCCA 417
Db 62 SerGluLysSerAspValGluGluAsnAsnAspHisAspAla----- 75
QY 418 CCCACAGCAGCAGCTGGCGCTGCTCCCTGGCGGCGAGTGGGAGGAGCAG----- 465
Db 76 ---ProLysGlnGluValLysLysThrArgLysThrSerSerArgLysThrLysAlaPro 94
QY 466 CCCAAGCGCACTGAGCGCGTGGAGCAGTTCCTGACCATTCGCGCGCGCGGAGGAGG 525
Db 95 ProLysLysThrThrLysGluAsnLys-----ValSerGlyLysLysArgThrArgLys 112
QY 526 AGCATGCTGTCTCCCTCGAGGATTTCTGCTGAGCCACGCTCTGCCCGCCACACAGCC 585
Db 113 ValSerLysLysGluPheAspAspGlu----- 121
QY 586 GAGCAGCGCTCCGAGGAGCGCGTGGAAAGCGCTTCTGAG-----ACCAGAAGCGGCC 539
Db 122 GluValSerGluAspLysAsnValAlaLysSerSerSerIleLeuArgArgSerSerArg 141
QY 640 CAGTCTGCTTCCACAGCTGTGAAGAACGACGACGCTCTTCTGAAAGGTGAAGAGGAGG 699
Db 142 LeuAlaAlaAlaAsnAlaAsnLysLysLysLysLysLysLysLysLysLysLysLys 156
QY 700 GATGACCACTGATGACACTCCGATAGTACAGCGAGTGCCTGACCTTGAAGAGCTTCAG 759
Db 157 ---GluAsnGlnAsnSerSerAspGluGluLysGluAspThrProLys----- 172
QY 760 AATCCCTTCCAGAGAGCGGAGCAGGAGCCACTGAGAGGCGCTCAAGAGGATCCAG 819
Db 173 ---LysSerLysArgGlnAlaLysLysLysLysLysLysLysLysLysLysLysLys 189
QY 820 AGTCGCTCGGAGAGAGCGCGGAGGAGGCTCCCGCGAGACTGTGGGCTCCAGGCC 879
Db 190 AlaGluGluGluGluSerGluLysGluSerProLysLysSerLysLysSerLysAla 209
QY 880 AGTGACACTGTGGAGGCGCTCTG---CCGAGTAGCAGAGGCGCCGAGNACGATCAGGG 936
Db 210 ThrLysProAlaLysLysValLysProThrLys----- 221
QY 937 GTTGTGCCAGGCTGTATTGATTTGCGCGAGCTCAACACAGCAGGTTTATGATTGC 996
Db 222 -----AlaAlaLysThrAlaLysLysLysLysLysLysLysLysLysLysLys 232
QY 997 GACATCAAGATGAGGAGCGCTGGAGACTTGGCCCGACCGAAGCCCTGAATGTGAGGTAC 1056
Db 233 -----GluAspAspGluProSerGluAspSerGluSerGlnGlyGlyGluGlu 250
QY 1057 GACCCCAACGCGCTGTATTGATTTGCGCGAGCTCAACACAGCAGGTTTATGATTGC 1116
Db 251 GlnGluGluAla-----LysSerGlnAspSerGlnLys----- 261
QY 1117 TGTGACCGCTGTGAAGATGTTTCATGCGGATTTGTGGGCAATTTCTGAGGCTCGAGG 1176
Db 262 -----AspSerGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 268
QY 1177 AGGCTTTTGGAAAGCAATGGGAGAGACTATATCTGCCCAACTGCACCATTTCTGCAAGT 1236
Db 269 -----GluGluAsnAspGlu-----GluMet 275

```





```
Db 6301 SerAspLysIleProGluLeuAlaLysGluValSerArgLysProGluGlu----- 6317
QY 655 GCTGTGAGGAGACGACGACGCTCTCTGAAAAGAGTGAAGAGGAGGATCACCACGATGAC 714
Db 6318 -----ProGluLysProGluGluProGluLysLeu-----AspGluSerAsp 6331
QY 715 ACCTCGATAGTACAGCGAGTGCCTGACCTTGAAGAGCTTCAGATCGCCTTCGACGG 774
Db 6332 LysLysProAspGluSerGluThrIleThrGlnLysPro-----LysArgArgLeu 6348
QY 775 AAGCGGGAACAGGAGCCCTCTGAGAGCCCTGAAAGGGATCCAGAGTCGCTGCCGGAAG 834
Db 6349 GlnLysLeuLysAlaProThrGluGluLysPheGluIleProLysValThrLeuArgLys 6368
QY 835 AAGCGCCGG-----GAGAGGGTCCCGCCGAGACTGTGGCTCCGAGGCC 879
Db 6369 ThrSerGlnLysValPheValProGluGluValThrLeuGluThrValGluLeuGluHis 6388
QY 880 AGTGACACTGTGAGGGCGCTCCGCCAGT----- 909
Db 6389 ValGluThrProGluValValGluProGluValGluLysArgValTrpSerProProPro 6408
QY 909 ----- 909
Db 6409 GluTyrGluThrTyrValProGluGluIleProGluLysGluProValGluLeuGluLys 6428
QY 910 -----AACGAGGAGCCCGAGACGATCAGGCGGTT 939
Db 6429 TyrGluLysTyrGluProProThrLysProLysAspGluGluGluAspLysGlyLys 6448
QY 940 GTGTCCAGGCTGGGAAGAT-----GACAGAGAGAGTAAGTTG----- 978
Db 6449 TyrGluArgLysProLysAspLysProGluProGluGluAspArgLysLeuLysLeuGly 6468
QY 979 GAGGGAAGCGGCTCAGGACATCAAGATGAGAGCGCTGGAGC----- 1023
Db 6469 LysGlyLys-----LeuArgProGluGluGluGluGluGluLysLysLys 6484
QY 1024 -----TTGGGCGGACGAG--CCTGAATGTGAGGTTAAGACCCCAAGCCCTGTAT 1074
Db 6485 ProProLysArgArgProGlySerProGluLysGluAlaGluLysProGlnLeuLysPro 6504
QY 1075 TGCATTTCGCGCCAGCTCTCAACAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGAA 1134
Db 6505 IleProLysLysLysProGlu----- 6511
QY 1135 TGGTTTCATGGCGATGTGTGTGGGCAATTCAGGCTCGAGGAGCGCTTTTGGAAAGGAAT 1194
Db 6512 -----GluGluLysLysAspLysValThrProLysPro 6522
QY 1195 GGGGAGAGCTATATCTGCCCAAACTGCCAAT-----CTGCAA 1233
Db 6523 GlyLysLys-----ProSerLysLysIleProAspArgGluProValGluLeuGlu 6539
QY 1234 GTCAGATGAGACTATCTCAGAAACCGGAGTACAGCAAGCT----- 1278
Db 6540 ProPheGluArgThrGluProGluIleLeuAspLysAspLysValProLeuGluLysPro 6559
QY 1279 -----AAATGGAGACTCGGAGTGTGTGATGTCGACCGATTTGACAGTATAGCAACA 1329
Db 6560 LeuLysProLysProGluProLysGluLysProSerIleGluProGluAlaProLysPro 6579
QY 1330 ATAGACAGAAGTCTAGCGAAGCAACAGGATTAAGGGTTAGATTGAGAAGCTGCAAT 1389
Db 6580 LeuGluLysProSerGluGluGluProLysLysLysLysLysGluArgPro--LysP 6599
QY 1390 CCAAGTGGCAAGAGAACTCAAGATCTCCAGCTGTGTATGAGGCGCTGGTCTCA 1449
Db 6599 roGluLysLysGluGluGluAlaGluValProSer----- 6510
QY 1450 AAATGTATTGGCCCGGGTGTGTCTGACGTGGCGCAGCCGACTCGGTGTACTCCAGTAAT 1509
```

```
Db 6611 -----Trp-----ArgGlyLysArg----- 6615
QY 1510 GACTGTATTCCTCAACACAGCCCGCAGCAATGAAGTTTCTTAAGCTCAGGTAAGACAG 1569
Db 6616 --LeuProProLysGluGluGluLysGluIleVal--LeuLysProPheLysLysGlu 6634
QY 1570 AAGCCA--AAGCCTAAAGAAAGATGAAGATGAAGCCAGAGAGCCAGTCTTCCGAAA 1626
Db 6635 LysProGluGluProLysProLysProLysProLysProGlyLysProTyrGluProGlu 6654
QY 1627 TGC----- 1629
Db 6655 IleProGluProGluLysThrProLeuGluProTyrThrLysProAspLysGluLysVal 6674
QY 1630 -----GGTGTCTCAGCAGGATTAATAATCTCTCTGTGCACAAGAGACGCTCCAGAA 1683
Db 6675 ProAspGlyValThrGluProValLysProGluAsp--GluGluLysProLysProGlu 6693
QY 1684 AAAAAAGAGACCCACAGTGAAGAGCAGTGGTGGTCCCTGCGCGAGTGAACACCTCGGG 1743
Db 6694 ---GluGluIleLysProLysLysGluArgIleLysProAspLysAspGlu----- 6709
QY 1744 AAGGAGCGCTGTGAGAGCAGCAGCGCTGTGG----- 1779
Db 6710 -----GluValGluThrProSerTrpArgGlyLysArgLeuProProLys 6724
QY 1780 GCGAGCGATCACATTAACAATGCAGTAAAGCCA-----GAAAAGAGCTGTGCTCCC 1830
Db 6725 GluGluAspLysGluGluIleThrLeuLysProPheLysLysGluLysProLysGluPro 6744
QY 1831 TCGCCGTCA 1839
Db 6745 LysProSer 6747

RESULT 12
Q8C969 PRELIMINARY; PRT; 443 AA.
ID Q8C969; AC Q8C969; DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
GN A630082K20RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AK042834; BAC31377.1; --
DR MGD; MGI:2443388; A630082K20RIK.
DR GO; GO:0003677; P:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR003347; TF_JmjC.
DR InterPro; IPR001965; ZnF_PHD.
DR Pfam; PF02373; JmjC; 1.
DR Pfam; PF00628; PHD; 1.
DR SMART; SM00558; JmjC; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS01186; EGF 2; 1.
DR PROSITE; PS01359; ZF_PHD 1; 1.
DR PROSITE; PSF0016; ZF_PHD_2; 1.
KW Hypothetical protein.
FT NON_TER 443
SQ SEQUENCE 443 AA; 49944 MW; C543F487717C4682 CRC64;
```

Alignment Scores: 0.000134 Length: 443  
Pred. No.: 187.00 Matches: 78  
Score: 187.00 Conservatives: 44  
Percent Similarity: 34.46% Mismatches: 118  
Best Local Similarity: 22.03% Indels: 114  
Query Match: 3.99% Gaps: 18  
DB: 11

US-09-787-016A-1 (1-2610) x Q8C969 (1-443)

QY 982 GGAAAGCGCGCTCAGGACATCAAGATGAGGAGCTGGAGACTTGGCGCGACCGAAGCCT 1041  
Db 15 GlyAlaAlaAlaAlaGlySerValSerAlaProGlyArgAlaSerAlaProPro 34  
QY 1042 GAATGTGAGGTTACGACCCCAAGCCCTGTATGCTATGCTGCGCAGCTCACAC--- 1098  
Db 35 -----ProProProValTyrCysValCysArgGlnProTyrAspVal 48  
QY 1099 AACAGGTTTATGATTTGCTGTGACCGCTGTGAGAGATGTTTCATGGCGATTTGGGCG 1158  
Db 49 AsnArgPheMetIleGluCysAspValCysArgGlnProTyrAspVal 68  
QY 1159 ATTTCTGAGGCTCGAGGAGGCTTTTGGAAAGGAATGGGGAAGCTATATCTGCCCAAC 1218  
Db 69 ValGluGluHisAlaValAspIleAspLeu-----TyrHisCysProAsp 84  
QY 1219 TGACACATCTCTCAAGTCAGGATGAGACTCATTGAGAGCGCGCATCAGCAGGAAGCT 1278  
Db 85 CysAlaAlaLeu-----HisGlySerSerLeuMetIleGlyArgArg 98  
QY 1279 AAATGGAGAGCTCGAGAT-----GCTGATGGCAGCGATTTGTACAACTATAGAA 1326  
Db 99 AsnTyrHisArgHisAspTyrThrGluValAspGlySerIleValGlnAlaGly 118  
QY 1327 ACA----- 1329  
Db 119 ThrArgAlaPheValIysGluLeuArgSerArgValPheProSerAlaAspGluIle 138  
QY 1330 -----ATAGACAGAGCTTAGCGAGTCTAGCGAGACCAAGG-----ATA 1362  
Db 139 ValIysMetHisGlySerGlnLeuThrGlnArgTyrLeuGluIysHisGlyPheAspVal 158  
QY 1363 AAGGTAGATTGAGAAAGCTGMAATCCAAAGTGGCAGAGAACTCAAGATCTTCCAG 1422  
Db 159 ProIleMetValProIysLeuAspAspLeuGly-----LeuArgLeuProSer 174  
QY 1423 CQTGTGATAGAGCGCTGCTGCTCAAAATGATTGGCCCGCGTCTGTCTCAGTGGCG 1482  
Db 175 ProAlaPheSerValMetAspValGluArgTyrValGly----- 187  
QY 1483 CAGCCGAGCTGGGTACTGCAAGTATGATCTCTCAACAGCGCGCAGCAATG 1542  
Db 188 ---GlyAspIysValIleAspValIleAsp---ValAlaArgGlnAlaAspSerIysMet 205  
QY 1543 AAGTTTCTAGCTCAGGTAAAGACAGACAGCAAGCTTAAGAAAGATGAAGATGAAG 1602  
Db 206 ThrLeuHisAsn-----TyrValIysTyrPheMetAsn 216  
QY 1603 CCAGAGAAGCCAGCTCTTCCGAAATGCGGTCTCAGGAGGTATTAATCTCTTCTGTG 1662  
Db 217 ProAspArgProIysVal-----LeuAsnValIleSerLeu 228  
QY 1663 CACAGAGACCGCTCCAGAAAAAAGACACACAGTGAAGAGCGAGTGGTCCCT 1722  
Db 229 -----GluPheSerAspThrIysMetSerGluLeuValGluValPro 242  
QY 1723 GCGCGAGTGAAGCACTCGGAGGAAGACGCTTTGTGAGAGCAGCAGCGCTGTGGCG 1782  
Db 243 AspileAlaArgIysLeu-----SerTyrVal 251  
QY 1783 AGCGATCAATTAATGAGTAAAGCAAGAAAGACTGCTGTCTCCCTCGCTCACTG 1842  
Db 1783 AGCGATCAATTAATGAGTAAAGCAAGAAAGACTGCTGTCTCCCTCGCTCACTG 1842

252 -----GluAsnTyr-----TipProAspAspSerValPheProIysProPheVal 266  
1843 TTGTATAAATGATG-----TATCAGCTAGGGTGGC 1875  
267 GlnIysTyrCysLeuMetGlyValGlnAspSerTyrThrAspPheHisIleAspPheGly 286  
1876 CTCCTGAGGAGCCCTCCGCTTCTTCTGGATAGCCATCCCTCG 1917  
287 -----GlyThrSerValTyrTyrHisValLeuTyr 296

RESULT 13  
Q8C980  
ID Q8C9E0 PRELIMINARY; PRT; 473 AA.  
AC Q8C9E0;  
DT 01-MAR-2003 (TRENBLrel. 23, Created)  
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE Hypothetical protein (Fragment).  
GN A630082K20RIK.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Thymus;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RL Nature 420:563-573(2002).";  
DR EMBL; AK042327; BAC31226.1; -;  
DR MGD; MGI:2443388; A630082K20RIK.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
DR InterPro; IPR006209; EGF like.  
DR InterPro; IPR003347; TF\_JmjC.  
DR InterPro; IPR001965; ZnF\_PHD.  
DR Pfam; PF02373; JmjC; 1.  
DR Pfam; PF00628; PHD; 1.  
DR SMART; SM00558; JmjC; 1.  
DR SMART; SM00249; PHD; 1.  
DR PROSITE; PS01186; EGF 2; 1.  
DR PROSITE; PS01359; ZF\_PHD\_1; 1.  
DR PROSITE; PS00016; ZF\_PHD\_2; 1.  
KW Hypothetical protein.  
FT NON TER 473 473  
SQ SEQUENCE 473 AA; 53409 MW; 94C94378609C7039 CRC64;

Alignment Scores:  
Pred. No.: 0.000136 Length: 473  
Score: 187.00 Matches: 78  
Percent Similarity: 34.46% Conservatives: 44  
Best Local Similarity: 22.03% Mismatches: 118  
Query Match: 3.99% Indels: 114  
DB: 11 Gaps: 18

US-09-787-016A-1 (1-2610) x Q8C9E0 (1-473)

QY 982 GGAAAGCGCGCTCAGGACATCAAGATGAGGAGCTGGAGACTTGGCGCGACCGAAGCCT 1041  
Db 15 GlyAlaAlaAlaAlaGlySerValSerAlaProGlyArgAlaSerAlaProPro 34  
QY 1042 GAATGTGAGGTTACGACCCCAAGCCCTGTATGCTATGCTGCGCAGCTCACAC--- 1098  
Db 35 -----ProProProValTyrCysValCysArgGlnProTyrAspVal 48  
QY 1099 AACAGGTTTATGATTTGCTGTGACCGCTGTGAGAGATGTTTCATGGCGATTTGGGCG 1158  
Db 49 AsnArgPheMetIleGluCysAspValCysArgGlnProTyrAspVal 68  
QY 1159 ATTTCTGAGGCTCGAGGAGGCTTTTGGAAAGGAATGGGGAAGCTATATCTGCCCAAC 1218



```

Db      69 ValGluGluHisHisValAspLeu-----TyrHisCysProAsp 84
QY      1219 TGCACCATTCCTGCAAGTGCAGCATGACATTCATTAGAAACCGCAGCATCAGCAGGAGCT 1278
Db      85 CysAlaAlaLeu-----HisGlySerSerLeuMetLysLysArgG 98
QY      1279 AATGAGACCTGGAGAT-----GCTGATGGCACCCGATTGTACAGTATAGGA 1326
Db      99 AsnTrpHisArgHisAspTyrThrGluValAspAspGlySerLysProValGlnAlaGly 118
QY      1327 ACA----- 1329
Db      119 ThrArgAlaPheValLysGluLeuArgSerArgValPheProSerAlaAspGluIleIle 138
QY      1330 -----ATAGACGAGAGCTTAGCGAAGCAACCAAGG-----ATA 1362
Db      139 ValLysMetHisGlySerGlnLeuThrGlnArgTyrLeuGluLysHisGlyPheAspVal 158
QY      1363 AAGGTTAGATTGAGAAAGCTGCAATCCAAAGTGGCAAGAGAAATCAAGATCTTCCAG 1422
Db      159 ProIleMetValProLysLeuAspAspLeuGly-----LeuArgLeuProSer 174
QY      1423 CCTGTGATAGAGCGCTGTGTGCTCAAAATGATTTGGCCCGCGTCTGTCACTGGCG 1482
Db      175 ProAlaPheSerValMetAspValGluArgTyrValGly----- 187
QY      1483 CAGCCGACCTGGTGTACTCGATGATGATCTCTCAACACGCGCAGCGACAATG 1542
Db      188 ---GlyAspLysValIleAspValIleAsp---ValAlaArgGlnAlaAspSerLysMet 205
QY      1543 AAGTTTCTAAGCTCAGGTAAAGAACAGCAAGCAAGCGCTAAAGAAAGATGAAGATCAAG 1602
Db      206 ThrLeuHisAsn-----TyrValLysTyrPheMetAsn 216
QY      1603 CCAGAGAACGCCAGCTCTTCGAAATCGGTGCTCAGGCGAGGTATTAATCTCTCTGTG 1662
Db      217 ProAspArgProLysVal-----LeuAsnValIleSerLeu 228
QY      1663 CACAGAGACCGCTCCAGAAAAGAGACAGCACAGTGAAGAGGCGAGTGTGTCCT 1722
Db      229 -----GluPheSerAspThrLysMetSerGlnLeuValGluValPro 242
QY      1723 GCGCGAGTGAAGCATCTCGGAGAGGACAGCTTGTGTGAGACGACGCGCGTCTGTGGCG 1782
Db      243 AspIleAlaArgLysLeu-----SerTrpVal 251
QY      1783 AGCGATCACAAATACAAATGAGTAAAGCCAGAAAGAGACTGCTGCTCGCGCTCACTG 1842
Db      252 -----GluAsnTyr-----TrpProAspAspSerValPheProLysProPheVal 266
QY      1843 TTGTATAAATCTANG-----TATCACTAGGGGTTGGC 1875
Db      267 GlnLysTyrCysLeuMetGlyValGlnAspSerTyrThrAspPheHisIleAspPheGly 286
QY      1876 CTCCTGGACCCCTCCCTCTCTTCTGATAGCCATCCCTGG 1917
Db      287 -----GlyThrSerValTrpTyrHisValLeuTrp 296

```

## RESULT 14

```

Q7SZE4 ID Q7SZE4 PRELIMINARY; PRT; 563 AA.
AC Q7SZE4;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
EN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=AB; TISSUE=Body;
RX MEDLINE=22388257; PubMed=12477932;
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altshuler S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards R.F., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A.C., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield V.S.,
RA Krzywinski M.I., Skalski U., Smalios D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AB; TISSUE=Body;
RX Strausberg R.;
RL Submitted (May-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC052918; AAH52918.1; --
KW Hypothetical protein.
SQ SEQUENCE 563 AA; 66587 MW; A0834B9C49192E83 CRC64;

```

```

Alignment Scores:
Pred. No.: 0.000179 Length: 563
Score: 185.50 Matches: 90
Percent Similarity: 35.64% Conservative: 49
Best Local Similarity: 23.08% Mismatches: 90
Query Match: 3.96% Indels: 161
DB: 17 Gaps: 17

```

US-09-787-016A-1 (1-2610) x Q7SZE4 (1-563)

```

QY 1006 GATGAGGACCTGGA---GACTTGGCGCCGACCGACCTGAATGTGAGGTTACGACCCC 1062
Db 9 AspGlnThrProGlyLeuAspAsnSerMetGluLysGlyGlu----- 22
QY 1063 AACGCC---CTGTATTGCTTTTCGCCAGCTCCACACACAGGTTATGATTTGCTGT 1119
Db 23 AsnAlaProLeuTyrCysIleCysArgLysSerAspIleAsnCysPheMetIleGlyCys 42
QY 1120 GACCGCTGCAGAAATGTTTCATGCGGATTGTGCGCATTTCTGAG----- 1167
Db 43 AspAsnCysAsnGlnTrpPheHisGlyHisCysIleAsnValThrGluLysMetAlaLys 62
QY 1167 ----- 1167
Db 63 AlaIleArgGluTrpTyrCysGlnGlnCysArgAlaArgAspProSerLeuSerIleArg 82
QY 1167 ----- 1167
Db 83 TyrArgLysLysAsnArgAspLysAspValGluProGluArgValGluLysArgSerSer 102
QY 1168 -----GCTCGAGGAGGCTTTTGGAAAGGATGG----- 1197
Db 103 ThrProGluTyrLysIleAspLysArgArgGlySerLysValLysArgSerAlaArgMet 122
QY 1198 -----GAAGACTATATCTGCGCAACTGCACCAT 1227
Db 123 CysGlyGluCysGluProCysThrArgThrGluAsp-----CysGlyHisCys-AspPh 140
QY 1228 CTCG-----AAGTGCAGGATGAG 1245
Db 140 eCysLysAspMetLysLysPheGlyGlyProAsnLysIleArgGlnLysCysArgLeuAr 160

```

```

QY 1245 A-----CTCATTCAGAAACCGCAGATCAGCAGGAAGCTAAATGGAGACCTGGAGATG-- 1297
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
160 gGlnCysValValArgAlaArgLysMetLeuArgValArgAspGluGluPheSerLeuArg 180
QY 1298 -----CTGATGCACCGA----- 1310
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
180 gGluArgLysAspAsnIleMet-HisArgAspArgGlyTyrSerAspAspTyrAspGluA 200
QY 1311 -----TTGTCAAGATATAGAACAAATAGA----- 1334
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
200 snAspMetAspLeuTyrGluHisTyrLysAspArgAsnAlaSerTrpGlySerGluAspA 220
QY 1335 -----GCAGAGAGTCT 1344
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
220 spAspGlyGlnLeuTyrSerProValProArgLysLysAlaIleLysValLysValL 240
QY 1345 AGCAGACACCAAGGAGTAAGGGTAGAATTGAGAAAGCTGCAATCAATCAAGTGCAGAAAG 1404
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
240 ysArgArgAspLysLysPheAspLysLysLysGluSerArgHisLysGlnLysGlnL 260
QY 1405 AAATCAAGAT--CTTCCAGCCTGTGATAGA-----GGCGCCTGTGCTGCC 1446
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
260 yHisArgAspArgLeuArgHisSerAspArgThrAspGlyArgHisGlyGly-AspThr 279
QY 1447 TCAAAATGATTGGCCCGCGTGTGCTGCAGTGGCGCAGCCGACTCGGTGTACTGCAGT 1506
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
280 GlnGlnCysLeuGlyProAsnCysIleGluProAlaArgProAsnSerLysTyrCysSer 299
QY 1507 AATGACTGTATCTCAACACCGCGCAGCGCAGCAATGAGTTTCTAAGCTCAGGTAAGAA 1566
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
300 GluAspCysGlyMetLysLeuAlaAsnArgIle-----TyrGlu 313
QY 1567 CAGAGCCAAAGCCTAAGAAAGATGAAGATGAAGCCAGAGACCCAGTCTTCCGAAA 1626
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
314 ValLeuProGlnArgIleGlnGlnTrpGlnGlnSerPro----- 326
QY 1627 TGGCGTGCT-----CAGGCGAGTATTAAATCTTCTGTGCAACAGACAGCTCCA 1680
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
327 CysIleAlaGluGlnGlyLysLysGlnLeuGluArgIleArg-ArgGluGlnGlnAl 346
QY 1681 GAAAAAAGAGA---CCACAGTGAAGAGCAGCTGTGTGCTCGCGAGTGAAGCA 1737
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
346 alaArgMetArgLeuAlaGluMetGluArgArgPheHisGlnLeuGluGlyIleIleAl 366
QY 1738 CTCGGGAAGGAGCAGCTGTG 1759
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
366 aLysAlaLysGlnLeuVal 373

RESULT 15
O74508 PRELIMINARY; PRT; 424 AA.
AC O74508;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE SPCC594.05c.
GN SPCC594.05c.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972h-;
RA Rieger M., Lyne M., Rajandream M.A., Barrell B.G.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL031523; CAA20664.1; -
DR PIR; T41449; T41449
DR GeneDB; SPombe; SPCC594.05c; -
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.

```

```

DR InterPro: IPR001965; Znf_PHD.
DR Pfam: PF00628; PHD; 1
DR SMART: SM00249; PHD; 1
DR PROSITE: PS01359; ZF_PHD_1; 1.
DR PROSITE: PS50016; ZF_PHD_2; 1.
SQ SEQUENCE 424 AA; 48483 MW; 2963605C3DFCC0B9 CRC64;

Alignment Scores:
Pred. No.: 0.000183 Length: 424
Score: 185.00 Matches: 96
Percent Similarity: 35.05% Conservative: 47
Best Local Similarity: 23.53% Mismatches: 148
Query Match: 3.95% Indels: 117
DB: 3 Gaps: 15

US-09-787-016A-1 (1-2610) x O74508 (1-424)
QY 535 GTCTCCCTGGAGGATTCGTGTGAGCCCGCAGCTCTCCCGCCGACACAGAC-----GCCGAG 588
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
20 ValLysPheGluAspSerAsnArgGlyThr-----IleThrAspPheHisIleGlu 36
QY 589 ACAGGCTCCGAGGGCGAGGTGGAAAGCGCTTCTGAGACCGAGAGCGGCCCGCCAGTCTGCT 648
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
37 ThrAlaAsnAsnGluGluGluLysAspAlaAsnValIleLeuAsn----- 51
QY 649 TCCACAGCTGTGAAGAGCAGCAGCAGCTCTTCTGAAAGCTGAAAGGGGATGACACAC 708
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
52 -----LysSerValLysMetGluValGluGluValAsnGly-----His 64
QY 709 GATGACACCTCGATAGTGCAGCAGCGATGCCCTTGAAGAGAGCTTCAGAAATCGCCTT 768
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
65 ValAspSerSerSerThrGluThrAsp-----Ile 74
QY 769 CCGAGGAACGGGAACAGGAGGAGCCACTGAGAGCGCCCTGAAAGGGATCCAGAGTCCCTG 828
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
75 GluMetGlnValIleGlnGlnProThrIleProLysLysProValSerAlaHisArg 94
QY 829 CGGAAGAACGGCGGAGGAGGAGTCCCGCGCAGACTGTGGGCTCCGAGGCCAGTGCACACT 888
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
95 ArgGlyProArgLysHisArgLysAsnAla----- 104
QY 889 GTGGAGGCGCTCTGCTCCCAAGTAGAGAGGAGCCCGAGAGCATCAGGGGGTTGTGTCAG 948
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
105 -----AsnSerGlnLeuAsnLeuSerThr 112
QY 949 GCTGGGAAGATGACAGAGAGAGTAAGTTGGAGGAAAGCGGCTCAGGACATCAAGAT 1008
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
113 Ala----- 113
QY 1009 GAGGAGCTGGAGACTTGGCGCCGACGAGCCTGAATGTGAGGGTTACGACCCCAAGCC 1068
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
114 -----AspHisGlnArgPro----- 118
QY 1069 CTGTATTGATTTGCGCCGAGCTCTACACACAGGTTTATGATTGCTGTGACCGCTGT 1128
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
119 LeuTyrCysIleCysGlnLysProAspAspGlySerTrpMetLeuGlyCysAspGlyCys 138
QY 1129 GAAGAAATGGTTTCATGCGCATTTGTGGGCATTTCTGAGGCTCGAGGAGGCTTTTGGAA 1188
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
139 GluAspTrpPheHisGlyThrCysValAsnIleProGluSerTyrAsnAspLeuThrVal 158
QY 1189 AGCAATGGGGAAGACTATATCTGCCAAACTGCACCATTTCTGCAAGTCGAGGATGAGACT 1248
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
159 Gln-----TyrPheCysProLysCysThr-----GluGluGly 169
QY 1249 CATTCAGAAACGGCAGATCAGCAGGAGCACTAAATGGAGA-----CCTGGAGAT 1296
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
170 LysGlyIleThrThrTrpLysArgLysCysArgLeuArgGluCysSerAsnProThrArg 189
QY 1297 GCTGATGGCCCGATTGTACAGTATAGGAACAATAGACAGAGTCTACGAGACCAA 1356
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
190 ProAsnSerAsnTyrCys-----SerAspLysHis 199

```

```
QY 1357 GGGATAAAGGGTAGAATTGAGAAA-----GCTGCAAAATCCAGTGGCAGAG 1404
Db |||:|:| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
200 GlyValAspPhePheArgGluLysValLysLeuSerThrValGluProSerAlaIleLys 219
QY 1405 AAACCTCAAGATCTTCAGGCTGTGATAGAGGCGCCTGTGTCTCAAAATGTATTGCCCC 1464
Db |||:|:| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
220 AsnLeuValLeuPheAlaLysLysArgGluPheGlnAsnLeuGlyThrValGlyPro 239
QY 1465 GGGTCTGTACAGTGGCGCAGCGGACTCGGTGTACTGCACTAATGACTGTATCCTCAA 1524
Db |||:|:| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
240 ThrLeuProSerGlnValProGluValValTyr-----AsnPheGluIleGluGlu 257
QY 1525 CACGCGCAGCGCAATGAAGTTTCTAGCTCAGGTAAAGAACAGAGCCAAAGCCTAAA 1584
Db |||:|:| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
258 AlaAsnArgLeuAsnAlaGluIleValGlnLeuAsnLysGluLysGluValAlaSerAsn 277
QY 1585 GAAAGATGAAGATGAAGCCAGAGCCCGAGTCTTCGAAATGCGGTCTC----- 1636
Db |||:|:| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
278 LysLysIlePheLeuGlnLeuIleLysAspSer-SerArgArgAlaValLeuAlaTyrIy 297
QY 1637 ---AGGCAGGTATTAAATCTCTCTGTGTGCACAGAGACCAGTCCAGAAAAAAGAG 1692
Db |||:|:| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
297 sGluArgGluGlyIleLysLysAspLeuCysglyPheAspSerArgLeuLeuPheAsnGl 317
QY 1693 ACCACAGTCAGAGAGCGACTGG 1714
Db : ||| |||
317 nGlnGlnMetAsnGluLeuTyr 324
```

Search completed: April 28, 2004, 11:00:23  
Job time : 200.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 28, 2004, 10:46:57 ; Search time 43 seconds  
(without alignments)

11677.204 Million cell updates/sec

Title: US-09-787-016A-1

Perfect score: 4689

Sequence: 1 ctccgtggcgcgcgcgcac.....actcttaagatcatactctg 2610

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp  
-O=/cgn2.1/USPTO spool\_p/US09787016/runat 28042004 094329 26935/app\_query.fasta\_1.2759  
-DB=PIR\_78 -OPMT=fastan -SUFFIX=xpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -WATFLX=blosum62 -TRANS=human40.cdi -LIST=45  
-LOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09787016 @CEN 1.1.66 @runat 28042004 094329 26935 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR\_78.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	193	4.1	488	2 I46014	Cylicin II - bovin
2	189.5	4.0	706	2 A45990	junctional sarcopl
3	185	3.9	424	2 T41449	probable phd finger
4	182	3.9	699	2 I38073	nucleolar phosphop
5	181	3.9	5327	2 T13564	microtubule-associ
6	180.5	3.8	689	2 JC5862	hepatoma-derived g
7	180	3.8	606	2 A43427	neurofilament trip
8	176.5	3.8	1110	2 I51116	NF-180 - sea lamp
9	174	3.7	734	2 B42680	nucleolus-cytoplas
10	170.5	3.6	510	2 T21430	hypothetical prote
11	169.5	3.6	990	2 I51618	nucleolar phosphop
12	169.5	3.6	1684	2 JMW057	gravin - human
13	169	3.6	723	2 S68191	tridinin - human
14	167.5	3.6	1087	1 QFM5H	neurofilament trip

# ALIGNMENTS

## RESULT 1

### I46014

Cylicin II - bovine

C:Species: Bos primigenius taurus (cattle)

C:Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 05-Nov-1999

C:Accession: I46014; S52772

R:Hess, H.; Heid, H.; Zimbelmann, R.; Franke, W.W.

Exp. Cell Res. 218, 174-182, 1995

A:Title: The protein complexity of the cytoskeleton of bovine and human sperm heads: the

A:Reference number: 137271; PMID:95255491; PMID:7737358

A:Accession: I46014

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-488 <HES>

A:Cross-references: EMBL:246789; NID:g757753; PIDN:CAA86753.1; PID:g757754

Alignment Scores:

Pred. No.: 8.98e-05 Length: 488

Score: 193.00 Matches: 117

Percent Similarity: 32.99% Conservative: 73

Best Local Similarity: 20.31% Mismatches: 204

Query Match: 4.12% Indels: 182

DB: 2 Gaps: 19

US-09-787-016A-1 (1-2610) x I46014 (1-488)

QY 316 CCCACCAGCAAGAGTTCAGGAAACATGGGTTTTCAGGAGGACCACCTATGCCAAGCGA 375

Db 19 ProValSerGluLeuSerLysSerTrpAsn----- 29

QY 376 GAGGCGCGAGGGGACCGCGGAGGCTGACCCACTGGAGCGCCGCCACACAGCAGCAGCTG 435

Db 30 -----GlnGlnHisPhe 33

QY 436 GGCCTGTCCCTCGGCGGCGAGTGGGAGGAGCCCAAGCGCAGCTGAGCGCGTGGAGCAGTTTC 495

Db 34 AlaLeuValPheProLysProArgProGlyLysArg----- 46

proliferation pote

hypothetical prote

C-terminal domain-

ribonuclease E P22

hypothetical prote

hypothetical prote

latent transforman

neatin - rat

protein pV100 [imp

hypothetical prote

antigen containing

recombination repa

hypothetical prote

hypothetical prote

hypothetical prote

calpastatin, cardi

TCOP1 protein - mo

tastin human

hypothetical prote

X-linked nuclear p

microtubule-associ

GTL protein - mous

neurofilament trip

juncnii protein -

Krox-20 protein -

hypothetical prote

nucleic acid bindi

RNase E [imported]

RNase E [similarit

probable transcrip

neatin - human



Db 138 ProLeuLysArgLysAspIleHisLysGluLysIleGluLysGlnGluLysProGluArg 157  
QY 811 GCGATCCAGATCGCTG-----CGGAAGAAGCGCGGAGGAGGCTCCGCGGAG 861  
Db 158 LysIleProThrLysValValHisLysGluLysGluLysGluLysValLysGlu 177  
QY 862 ACTGTGGCTCCGAGCCAGTACACTGTGCGAGGCGCTCTCCGCGAGTAAAGCAGAGCC 921  
Db 178 LysGluLysProGluLysAlaThrHisLysGluLysLeuLysLysLysPro 197  
QY 922 GAGAACGATCAGGGGTGTGTCAGGCTGCGAAGATGACAGAGAGTAGTGTGGAG 981  
Db 198 GluThr-----LysThrValThrLysGluLysIleAlaArgThrLysGluLysIleGlu 216  
QY 982 GGAAGCGGCTCAGGACATCAAGATGAGAGCGCTGGAGACTGGCGCA-----1032  
Db 217 GluLysThrLysLysGluValLysGlyValLysGlnGluLysValLysGlnThrValAla 236  
QY 1033 -----CGGAGCCT---GAATGTCAGGTTACGACCCCAAC 1065  
Db 237 LysAlaLysGluValGlnLysThrProLysGluLysGluLysSerLysGluThrAla 256  
QY 1066 GCGCTGTATTGCAATTTGCCCGCCAGCTCAACAACAGAGTTTATGATTTGCTGTGACCGC 1125  
Db 257 Ala-----ValSerLysGlnGluLysAspGlnTyrAlaPheCysArgTyrMet 273  
QY 1126 TGTGAAGATGCTTATGCGGATGTGTGGCGATTTCTGAGCTCGAGGAGGCTTTG 1185  
Db 274 IleAspIlePheValHisGlyAsp-----Leu 282  
QY 1186 GAAAGGAATCGGAGACTATCTGCCCCAACTGCACCACTTCTGCAAGTGCAGGATGAG 1245  
Db 283 LysProGluLysProAlaIleProProSerProThrGlnAlaSerArgPro 302  
QY 1246 ACTCATTCA---GAAACGGCAGATCAGCAGAGAGCTTAATGGAGACTCGAGATGCTGAT 1302  
Db 303 ThrProAlaLeuProThrProGluLysGluLysGluLysLysLysAlaGlu-----320  
QY 1303 GGCACCGATTGTACAAGTATAGGAACAATAGACAGAGCTAGCAGAGCTAGCAAGAGGATA 1362  
Db 321 -----LysLysValThrThrGluThrLysLysLysAlaGluLysGluAspAla 336  
QY 1363 AAGGCTAGAAATTCAGAAAGCTGCAAGTCCAAAGTGGCAAGAGAACTCAAGATCTTCCAG 1422  
Db 337 LysLysLysSerGluLysGluThrAspIleAspMetLysLysLys-----351  
QY 1423 CTTGTGATAGAGGCGCTGCTGCTCAAAATGATTGGCCCGGCTGCTCAGCTGGCG 1482  
Db 352 -----GluProGluLysSerProAspThrLysProGluThrValLysValThr 367  
QY 1483 CAGCCCGACTCGGTACTGCAATGATGACTGTATCTCTCAACACGCGCGAGCAGCAATG 1542  
Db 368 Thr-----GlnAlaAlaThrLys 373  
QY 1543 -----AGTTTCTTAAGCTAGGTAAAGAACAGAGCCA-----AAGCTTAA 1584  
Db 374 LysAspGluLysLysGluAspSerLysLysAlaLysLysProAlaGluGlnProLys 393  
QY 1585 GAAAGATGAGATGAGAGCAGAGAGCCAGCTCTCCGAAATCGCGTGT-----1635  
Db 394 GlyLysLysGlnGluLysLysGluLysHisGluGluProAlaLysSerThrLysLysGlu 413  
QY 1636 -----CAGCGAGGTATTAAATC-----1653  
Db 414 HisAlaAlaProSerGluLysGlnAlaLysAlaLysIleGluArgLysGluGluValSer 433  
QY 1654 TCTTCTGTGACAGAGACCGCTCCAGAAAAAAGAGAGCCAGCTGGAAGAGGCGAGTG 1713  
Db 434 AlaAlaSerThrLysLysAlaValProAlaLysLysGluLysThrLysThrVal 453  
QY 1714 GTGGTCCCTCGCGGAGTGAAGCACTCGGGAAGGAAGCAGCTGT-----1758

Db 454 GluGlnGluThrArgLysGluLysProGluLysIleSerSerValLeuLysAspLysGlu 473  
QY 1759 -----GAGAGCAGCAGCGCTGCTGGCGAGCGATCACAATTACATGCA 1803  
Db 474 LeuThrLysGluLysValProAlaSerLeuLysGluLysGlySerGluThr 493  
QY 1804 GTAAGCCAGAAAGACTGCTCTCCCTCGCGCTCACTG 1842  
Db 494 LysLysAspGluLysThrSerLysProGluProGlnIle 506  
RESULT 3  
T41449  
probable phd finger transcription regulator - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 07-Dec-1999  
A:Accession: T41449  
R:Rieger, M.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, September 1998  
A:Reference number: Z21994  
A:Accession: T41449  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-424 xRIS>  
A:Cross-references: EMBL:AL031523; PIDN:CAA20664.1; GSPDB:GN00068; SPDB:SPCC594.05C  
A:Experimental source: strain 972b; cosmid c594  
C:Genetics:  
A:Gene: SPDB:SPCC594.05C  
A:Map position: 3  
Alignment Scores:  
Pred. No.: 0.000293 Length: 424  
Score: 185.00 Matches: 96  
Percent Similarity: 35.05% Conservative: 47  
Best Local Similarity: 23.53% Mismatches: 148  
Query Match: 3.95% Indels: 117  
DB: 2 Gaps: 15  
US-09-787-016A-1 (1-2610) x T41449 (1-424)  
QY 535 GTCTCCTCGAGATCTCTGTGAGCCCACTGCTGCTCCCGCCACAGAC-----GCCGAG 588  
Db 20 ValLysPheGluAspSerAsnArgGlyThr-----IleThrAspPheHisIleGlu 36  
QY 589 ACAGCTCCGAGGCGAGCGTGAAGAGCGCTTCTGAGCCAGAGAGCGGCCCGCTGCT 648  
Db 37 ThrAlaAsnAsnGluLysGluLysAspAlaAsnValIleLeuAsn-----51  
QY 649 TCCACAGCTGTGGAAGAACGACGACGCTCTTCTGAAAGGTGAAAGGAGGATGACAC 708  
Db 52 -----LysSerValLysMetGluValGluLysValAsnGly-----His 64  
QY 709 GATGACACTCCGATAGTACAGCGATGCGCTGACCTTGAAAGAGCTTCAGATCGCCTT 768  
Db 65 ValAspSerSerThrGluThrAsp-----Ile 74  
QY 769 CGCAGGAACGGGAGACGAGCCCACTGAGAGCGCCCTGAAAGGATCCAGATCGCCTG 828  
Db 75 GluMetGlnValIleGlnGlnProThrIleProLysLysProValSerAlaHisArg 94  
QY 829 CGGAAGAAGCGCGGAGGAGGCTCCCGCCGAGACTGTGTGGCTCCGAGCCAGTGCACCT 888  
Db 95 ArgGlyProArgLysHisArgGlyAsnAla-----104  
QY 889 GTGAGGCGGTCTCTCCCGAGTAAGCAGGAGCCCGAGAACGATCAGGGGTGTGTGCCAG 948  
Db 105 -----AsnSerGlnLeuAsnLeuSerThr 112  
QY 949 GCTGGGAAAGATGACAGAGAGAGTAAAGTTGGAGGAAAGCGCGCTCAGGACATCAAGAT 1008  
Db 113 Ala-----113  
QY 1009 GAGGAGCTTGGAGACTTGGGCGCAGCGAAGCCTGATGTGAGGGTTACGACCCCAACGCC 1068



```

114 -----AspHisGlnArgPro----- 118
1069 CTGTATTGCAATTTGCCCGCAGCTCACACACAGGTTTATGATTGCTGTGACGCTGT 1128
119 LeuTyrCysIleCysGlnLysProAspAspGlySerTrpMetLeuGlyCysAspGlyCys 138
1129 GAAGATGTTTCAAGCGAATGTTGTGGCATTCTGAGGCTCGAGGAGGCTTTTGAA 1188
139 GluAspTrpHeHisGlyThrCysValAsnIleProGluSerTyrAsnAspLeuThrVal 158
1189 AGGAATGGGAGAGACTATATCTGCCAAACTGCACCACTTCTCAAGTGCAGGATGAGACT 1248
159 Gln-----TyrPheCysProLysCysThr-----GluGluGly 169
1249 CATTCAGAAACCGCAGATCAGCAGGAGCTAATGGAGA-----CCTGGAGAT 1296
170 LysGlyIleThrThrTrpLysArgLysCysArgLeuArgLysCysSerAsnProThrArg 189
1297 GCTGATGGCAGCATGTTGTACAGTATAGGAACAATAGACAGCAAGTCTAGCGAAGCAAA 1356
190 ProAsnSerAsnTyrCys-----SerAspLysHis 199
1357 GGATAAAGGTTAGATTGAGAAA-----GCTGCAAAATCCAGTGGCGAAGAG 1404
200 GlyValAspPheArgGluLysValLysLeuSerThrValGluProSerAlaIleLys 219
1405 AAATCAAGATCTTCGAGCTGTGATAGAGCGCTGTGCTCAAAATGTTATGGCC 1464
220 AsnLeuValLeuPheAlaLysLysArgGluGluPheGlnAsnLeuGlyThrValGlyPro 239
1465 GGTGCTGTCACTGCGCGCAGCGCCTGGTGTACTGCGAGTAAGTACTGTATCTCTCAA 1524
240 ThrLeuProSerGlnValProProGluValValTyr-----AsnPheGluIleGluGlu 257
1525 CAGCGCGCAGCGCAATGAAGTTTCTAAGTCAAGTAAAGACAGCAGCAAGCAAGCTTAA 1584
258 AlaAsnArgLeuAsnAlaGluIleValGlnLeuAsnLysGluLysGluValAlaSerAsn 277
1585 GAAATCAAGATGAAGCAGCAGCAGCAGCTTCCGAATTCGCGTCTC----- 1636
278 LysLysIlePheLeuGluIleLysLysAspSer-SerArgAlaValLeuAlaTyrIly 297
1637 ----AGGCGAGTAAATCTCTTCTGTGCAAGACAGCAGCTCCAGAAAAAAGAG 1692
297 sGluArgGluGlyIleLysLysAspLeuLysGlyPheAspSerArgLeuLeuPheAsnGly 317
1693 ACCACAGTGAAGAGGCACTGG 1714
317 nGlnGlnMetAsnGluLeuTrp 324

RESULT 4
I38073
nucleolar phosphoprotein p130 - human
C;Species: Homo sapiens (man)
C;Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 24-Sep-1999
C;Accession: I38073; S52292
R;Pal, C.Y.; Chen, H.K.; Sheu, H.L.; Yeh, N.H.
J. Cell Sci. 108, 1911-1920, 1995
A;Title: Cell-cycle-dependent alterations of a highly phosphorylated nucleolar protein p
A;Reference number: I38073; MUID:95386590; PMID:7657714
A;Accession: I38073
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-699 <RES>
A;Cross-references: EMBL:Z34289; NID:9663007; PIDN:CAA84063.1; PID:9663008
C;Superfamily: nucleolus-cytoplasm shuttle phosphoprotein
C;Keywords: phosphoprotein

Alignment Scores:
Pred. No.: 0.000489 Length: 699
Score: 182.00 Matches: 117
Percent Similarity: 36.10% Conservative: 83
Best Local Similarity: 21.12% Mismatches: 213

```

```

Query Match: 3.88% Indels: 141
DB: 2 Gaps: 19
US-09-787-016a-1 (1-2610) x I38073 (1-699)
QY 268 GACGACAAAGCGACCCAGCAATGAGGAGGACCTAAGGCCATCAAAACCCACAGCAAA 327
DB 136 AspAspGluAspGlnLysLysGlnProValGlnLysGlyValLysProGlnAlaLys 155
QY 328 GAGTTTCAGGAAAAATGGGGTTTTCGAAGGACCACTATGCCAAGCAGGAGGCGCAGGG 387
DB 156 AlaAlaLysAlaPro-----ProLysLysAlaLysSerSer 167
QY 388 GACGCGAGGCGTGAC---CCACTGGAGCGCCACCCACAGCAGCAGCTGGCGCTGTCC 444
DB 168 AspSerAspSerAspSerSerSerGluAspLupProLysAsnGln----- 183
QY 445 CTGCGGCGCAGTGGAGAGGACCCCAAGCGCACTGAGCGCTGGAGCAGTTCCTGACCACT 504
DB 184 -----LysProLysIleThr----- 188
QY 505 GCGGCGCGCGCGGAGGAGGAGCATGCTCTCTCCCTGGAGGATTCTGTGAGCCCCACG 564
DB 189 -----ProValThrValLysAlaGlnThrLysAlaPro 199
QY 565 TCCTGCCCCCGCACAGACGCGGAGCAGCTCCGAGGCGCGGTGGAAGCGCTTCTCTGAG 624
DB 200 ProLysProAlaArgAlaAlaProLysIleAlaAsnGlyLysAlaAlaSerSerSer 219
QY 625 ACCAGAGCGCGCCCGCAGTCTCTCCACACTGTGAGGAGAACACACACCTCTTCTCGAA 684
DB 220 SerSerSerSerSerSerSerSerSerSerGluGluGluLysAlaAlaAlaThrPro 239
QY 685 AAG-----GTCAAGAGGAGGAGTAC 705
DB 240 LysLysThrValProLysLysGlnValValAlaLysAlaProValLysAlaAlaThrThr 259
QY 706 CACGATGACACTCGATGATGACAGCGATGCGCTGACCTTGAAGAGCTTCAAGATCGC 765
DB 260 ProThrArgLysSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 279
QY 766 CTTCGACAGAAAGCGGAAACAGAGGCGCC-----ACTGAGAGG 801
DB 280 LysProMetLysAsnLysProGlyProTyrSerTyrAlaProProProSerAlaProPro 299
QY 802 CCGCTGAAAGGATCCAGATGCGCTGCGGAGAAAGCCCGGAGGAGGCTCCCGCCGAG 861
DB 300 ProLysLysSerLeuGlyThrGlnProProLysLysAlaValGluLysGlnGlnProVal 319
QY 862 ACTGTGGGCTCCGAGCGCAGTACACTGTGGAGGCGCTCTGCCCAAGTAAGCAGAGCGCC 921
DB 320 GluSerSerGluAspSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 339
QY 922 GAGAACGATCAGGGGGTGTGTCTCCAGGCT-----GGGAAAGATGAC 963
DB 340 ProThr-----LysAlaValValSerLysAlaThrThrLysProProProAlaLysAla 358
QY 964 AGAGAGGTAGTTCGAGGGAAGCGGCTCAGGACATCAAGATGAGGAG----- 1014
DB 359 AlaGluSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 378
QY 1015 -----CCTGAGACTTGGGCGCGCAGGAGCTGTAATGTGAGGGTTAGACCCCAAGCGC 1068
DB 379 LysProAlaGly-ThrThrLysAsnSerSerSerSerSerSerSerSerSerSerSer 398
QY 1069 CTGTATTGATTTGCGCGCAGCTCACACACAGCTTTATGATTGCTGTGACCGCTGT 1128
DB 398 oAlaValLysProAlaAlaAlaAlaProLysGlnProValGly----- 411
QY 1129 GAAGATGTTTTCATGGCGATTGTGGGCAATTTCTGAGGCTCGAGGAGGCTTTTGAA 1188
DB 412 -----GlyGlyGln-LysLeuLeuThr 418

```



Db	3250	LysSerProLeuProSerLysGluAlaSerArgProThrSerVal	---	3264
Qy	1087	CAGCCTCACAACAACAGGTTTATGATTGCTGTGACCGCTGTGAAGAATGGTTTCATGGC	1146	
Db	3264	-----	---	3264
Qy	1147	GATTGTGGGCATTTCTGAGGCTCGAGGAGGCGCTTTTGAAGGANTGGGAAGACTAT	1206	
Db	3265	---AlaLysSerValLysAspGluAlaGluLysSerLysGluSerSerArgAspSer	3283	
Qy	1207	ATTCGC-----CCAAACTGCACCATTCGT	1230	
Db	3284	ValAlaGluLysSerProLeuAlaSerLysGluAlaSerArgProAlaSerValAlaGlu	3303	
Qy	1231	CAAGTCGAGGATGAGCT-----CATTCAGAAACGGCAGAT	1266	
Db	3304	SerValGlnAspGluAlaGluLysSerLysGluLuserArgArgGluSerValAlaGlu	3323	
Qy	1267	CAGCAGGAAGCTAAATGGAGCTCGGAGATGCTGATGCCACCGATTGTCAAGATPATAGGA	1326	
Db	3324	LysSerProLeuAlaThrLysGluAlaSerArgProAlaSerValAlaGluSerIleLys	3343	
Qy	1327	ACAATAGCAGAGAGTCTAGCGAAGACCAAGGGATAAGGTAGAAATTGAGAAAGCT---	1383	
Db	3344	AspGluAlaGluLysSerLysGluLuserArgArgGluSerValAlaGluLysSerPro	3363	
Qy	1384	---GCAAAATCCAACTGGCCAAAGAAACTCAAGATCTTCAGCCTGTGATAGAGCGGCCT	1440	
Db	3364	LeuAlaSerLysGluAlaSerArgProThrSerValAlaGluSerValLysAspGluAla	3383	
Qy	1441	GGTCCTCAAAATGATTGGCCCGGGTGCTGTCACGTGGCGAGCCGCACTCGGTCTAC	1500	
Db	3384	GluLysSerLys-----GluGluSerSerArgAspSerValAla	3396	
Qy	1501	TGCAGTAATGACTATCTCTCAACACGCGCAGCGCAATGAAGTTCTTAAGCTCAGGT	1560	
Db	3397	GluLysSerProLeuAlaSerLysGluAlaSerArgProAlaSerValAlaGluSerVal	3416	
Qy	1561	AAAGACACAGAACCAAGCCCTAAAGAAAGATGAAGATGAAG-----CCAGAGAGAGCCC	1614	
Db	3417	GlnAspGluAlaGluLysSerLysGluLuserArgArgGluSerValAlaGluLysSer	3436	
Qy	1615	AGTCTTCGAAATCGGTGCTCAGGCAGGTATTAAATCTCTCTGTCACACAGAGACCA	1674	
Db	3437	ProLeuAlaSerLysGluAlaSerArgProAlaSerValAlaGluSerValLysAspAsp	3456	
Qy	1675	GCTCCAGAAAAAAGACACACACTGTAAGAAAGCAGGTGGTCCCTCGCGCGAGTGAA	1734	
Db	3457	AlaGluLysSerLysGluLuserArgArgGluSerValAla-----GluLysSerPro	3474	
Qy	1735	GCATCTGGGAGGAGCAGCTGTGTAGACGACGACCGCGTGTGGCGAGGAGCAC---	1791	
Db	3475	LeuAlaSerLysGluAlaSerArgProAlaSerValAlaGluSerValLysAspGluAla	3494	
Qy	1792	-----AATTCAATGCTAAAGCCAGAAAGACTGCTGCTCCCTCGCGC	1836	
Db	3495	GluLysSerLysGluLuserArgArgGluSerValAlaGluLysSerPro	3511	

## RESULT 6

hepatoma-derived growth factor-related protein 2 - mouse  
N:Alternate names: HRP-2  
C:Species: Mus musculus (house mouse)  
C:Date: 17-Nov-1997 #sequence\_revision 17-Nov-1997 #text\_change 05-Nov-1999  
C:Accession: JC5662  
P:Izumoto, Y.; Kuroda, T.; Harada, H.; Kishimoto, T.; Nakamura, H.  
Biochem. Biophys. Res. Commun. 239, 26-32, 1997  
A:Title: Hepatoma-derived growth factor belongs to a gene family in mice showing significant  
A:Reference number: JC5660; MUID:97445118; PMID:9399445  
A:Accession: JC5662  
A:Molecule type: mRNA  
A:Residues: 1-669 <IZU>  
A:Sequences: 1-669 <IZU>

C;Comment: This protein translocates to the nucleus and directly functions in mitogenic F;1-93/Region: hath #status predicted  
F;321-363/Region: nuclear location signal

Alignment Scores:	
Pred. No.:	Length:
Score:	669
180.50	Matches:
33.44%	Conservative:
20.86%	Mismatches:
3.85%	Indels:
2	Gaps:
2	DB:

Qy	220	AGGATTTCACGAGAAAAGTGTCCAAAGCTTTTCAGTGTGTGGACGAGGTATGGACGACGCAAAAGGC	279
Db	171	ArgArgAlaserSerAspLeuAspGlnAlaSerValSerProSerGluGluAspSerGlu	190
Qy	280	GACCCGAGCAATCAGGAGGACACCTTAAGGCCATCAAAACCCACAGACGAAAGATTCAGGAAA	339
Db	191	SerProSerGluSerGlu-----LysThrSerAspGlnAspPhe-----	203
Qy	340	ACATGGGGTTTTCGAAGCACCACTATCGCAACGAGGAGGCGCAGGGACGCGGAGGCT	399
Db	204	-----ThrProGluLeuLysThrAlaAla-----	211
Qy	400	GACCCACTGAGCGCGCACCCACACAGCAGCAGCTGGCGCTGTCCCTGGCGCGCAGTGGG	459
Db	212	-----ArgProProArgArgGlyProLeuGly-----	221
Qy	460	AGGCAGCCCAAGCGCACTGAGCGCGTGAGCACTTCTCACCATTGCGCGCGCCCGCGC	519
Db	222	ArgLysLysLysLysValProSerAlaSerAspSerAspSerLysAlaAspSerAspGly	241
Qy	520	AGGAGGAGCANGCTGTCTCCCTCGAGGATTCGTGTGACCCACAGTCTCTCCCGCCACCA	579
Db	242	AlaLysGluGluProValVal-----ThrAlaGlnProSer-----ProSerSer	256
Qy	580	GACGCCGAGACAGCTCCGAGGGCGAGCGTGGAAAGCGCTTCTCAGACACGAGAGCGGCC	639
Db	257	SerSerSerSerSerSerSerSerSerSerAspSerAspValSerValLysLysProPro	276
Qy	640	CAGTCTCCTCCACAGCTGTGAAG-----	663
Db	277	ArgGlyArgLysProAlaGluLysProProProLysProArgGlyArgArgProLysPro	296
Qy	664	GAACGACACGCTCTCTGTGAAAGGTGAAAGGAGGGATGACCCAGATGACACTCCGAT	723
Db	297	GluArgProProSerThr-----SerSerSerSerAsp	306
Qy	724	AGTCACAGCATGGCTTCGACTTGAAGAGCTTCAGAAAT-----	762
Db	307	SerAspSerAspSerGlyGluValAspArgLysSerGluThrLysArgArgAspGluGlu	326
Qy	763	-----CCCTTCGAGGAAGCGGACAGGAGCCACTGAGAGCGCCCTG	807
Db	327	ArgArgArgGluLeuGluAlaArgArgArgArgGluGluGluLeuArgArgLeu	346
Qy	808	AAAGGATCAGAGTCCCTCGCGGAAGACCGCGGAGGAGGCTCCCGCGAGACTGTG	867
Db	347	ArgGlu--GinGluA--GluGluLysGluArgArgLysGlu--ArgAlaGluArgGly	364
Qy	868	GGCTCCAGGCGCAGTGCACCTGTGGAGGGCGTCTGTGCCAGTAAGCAGAGAGCCGAGAAC	927
Db	365	GlySer-----SerGlyGluGluLeuGluAsp	373
Qy	928	GATCAGGGGTGTGTCCAGCTGGCGAAAGATGACAGAGAGATGAATGTGGAGGGAAG	987
Db	374	GluGluProVal-----LysLysArgSerArgLysAlaArgGlyArg	387
Qy	988	CGCGCTCAGGACATCAAGATGAGAGCCT---GGAGACTTGGCCCA-----	1032
Db	388	GlyThrProSerSerSerAspSerGluProGluGlyGluLeuGlyLysGluGlyLysLys	407

1033 QY -----CCGAAGCCTGAATGTGAGGGTTCAGACCCCAAGCCCTGTAATGCAATTTCCCGC 1086  
 408 Db LeuAlaLysSerGlnLeuProGlySerGluSer-----AlaArg 421  
 1087 QY CAGCCTCAACAACAGGTTTATGATTGCTGTGACCGCTGTGAAGAATGTTTCATGCC 1146  
 422 Db LysPro----- 423  
 1147 QY GATTGTGTGGGCATTCTTGAGCTCGAGGGAGCGCTTTTGGAAAGGAATGGGAAGACTAT 1206  
 424 Db -----GlyGlnLysGluLysArgGlyArg----- 431  
 1207 QY ATCTGCCCAACATGCACCATCTTCTCAAGTGCAGGTGAGACTCATTCAGAAACGGCAGAT 1266  
 432 Db -----Pro 432  
 1267 QY CAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGATTGTACAGTATAGGA 1326  
 433 Db AspGluLysProArgAlaArgProValLysValGluArgThr----- 446  
 1327 QY ACAATAGAGCAGGAAGTCTACGCAAGACCAAGGGATGAAGGTAGAAATTGAGAAGCTGCA 1386  
 447 Db -----ArgLysArgSerGluGlyLeuSerLeuGluArgLysGlyGluLysLys 463  
 1387 QY AATCCAGTGGCAAGAGAACTCAAG-----ATCTCCAGCCTGTG 1428  
 464 Db GluProSerValGluArgLeuGlnLysLeuHisSerGluLeuLysPheAlaLeuLys 483  
 1429 QY ATAGAGCGCTGTGCTCAAAATGATTATGGCCCC-----GGGTCTGTCTAC 1476  
 484 Db ValAspAsnProAspValArgLysCysLeuSerAlaLeuGluLeuGlyThrLeuGln 503  
 1477 QY GTGGCG-----CAGCCCGACTCGGTG----- 1497  
 504 Db ValThrSerGlnLeuGlnLysAsnThrAspValValAlaThrLeuLysLysIleArg 523  
 1498 QY ---TACTGCAATGATGCTGATCTCTCAACACGCGCA----- 1533  
 524 Db ArgTyrLysAlaAsnLysAspValMetAlaLysAlaAlaGluValTyrThrArgLeuLys 543  
 1534 QY -----GCGACAATGAAGTTCTTAAGCTCAGGTAAAGAA 1566  
 544 Db SerArgValLeuGlyProLysValGluAlaLeuGlnLysValAsnLysAlaGlyValaGlu 563  
 1567 QY CAGAGCCAAAGCTTAAGAAAGATCAAGTGAAGCCAGACAGAGCCAGTCTCCGAA 1626  
 564 Db LysGluArgAlaAspAsnGluLysLeuGluGlnPro----- 576  
 1627 QY TGGGTGCTCAGCAGGTATTAAATCTTCTCTGTGCAC-----AAGAGACAGCT 1677  
 577 Db ---GlyGluGlnAlaProArgGluLeuAlaGluAspGluProSerThrAspArgSerAla 595  
 1678 QY CCAGAAAAAAGAGACCACAGTGAAGAGCGAGTGTGTCTCCGCGGAGTGAAGCA 1737  
 596 Db ProValAsnGlyGluAlaThrSerGlnLysGlyLysAsnMetGluAspArgAlaGlnGlu 615  
 1738 QY CTGGGAAGCAAGCAGCTTGTAGAGCAGCACCGCGTGTGGCGAGCGATCAC----- 1791  
 616 Db AspGlyGlnAspSer-----GluAspGlyProArgGlySerSerGluLeuHis 633  
 1792 QY -----AATTCAATGCAGTAAAGCA-----GAA 1815  
 634 Db AspSerProArgAspAsnSerAspProAlaLysProGlyAsnGluArgGlnAspHisGlu 653  
 1816 QY AAGACTGCTCTCCCTCCGCTCA 1839  
 654 Db ArgThrArgLeuAlaSerGluSer 661

RESULT 7

A43427

neurofilament triplet H1 protein - rabbit (fragment)

C:Species: Oryctolagus cuniculus (domestic rabbit)

C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 10-Dec-1999

C:Accession: A43427

R:Soppet, D.R.; Beasley, L.L.; Willard, M.B.

J. Biol. Chem. 267, 17354-17361, 1992

A:Title: Evidence for unequal crossing over in the evolution of the neurofilament polype

A:Reference number: A43427; PMID:92381055; PMID:1512270

A:Accession: A43427

A:Molecule type: DNA

A:Residues: 1-606 &lt;SOP&gt;

A:Cross-references: GB:M94315; MID:g164990; PIDN:AAA57152.1; PID:g601930

A:Experimental source: brain

A:Notes: sequence extracted from NCBI backbone (NCBIN:112010, NCBI:P:112011)

C:Superfamily: neurofilament triplet H protein

C:Keywords: coiled coil; heterotrimer; intermediate filament; nerve; phosphoprotein

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
0.000647	180.00	606	106
Percent Similarity:	34.9%	Conservative:	77
Best Local Similarity:	20.2%	Mismatches:	211
Query Match:	3.84%	Indels:	129
DB:	2	Gaps:	19

US-09-787-016A-1 (1-2610) x A43427 (1-606)

QY 312 CAACCCACACCAAGACTTCAGAAACATGCGGTTTTCGAAGACCACTATCGCAA 371  
 Db 10 GlnThrGluGlnValThrGluGluVal-----ThrGluGluGluLysGlu 27  
 QY 372 GCGAGAGGCGCGAGGAGCGCGAGGCTGACCCATCGAGAGCGCCACCCACACA 425  
 Db 28 Ala-----LysGluGluGluGlyGlyGluGluGluGluAlaLysSerProThrGluGly 45  
 QY 426 GCAGCAGCTGGGCGCTCTCCCTGGCGGAGTGGAGGCGAGCCAGCGCCTGAGCGCT 485  
 Db 46 GlyAlaAlaSerProGluGluGluAlaLys----- 55  
 QY 486 GGAGCAGTTCTCTGACCATTCGCGCGCGCGAGGAGGAGCCTCTCTCTCCCTGA 545  
 Db 56 -----SerProAlaGluAlaLys-SerProValLys---G1 66  
 QY 546 GATTTCTGTGAGCCACGCTCTGCGCCCGCCACAGACGCGGAGACAGCTCCGAGGCGAG 605  
 Db 66 uGluAlaLysSerPro-----AlaGluAlaLysSerProAlaG1 79  
 QY 606 COTGAAAGCGCTTCTGAGACCAAGCGCGCCCGAGTCTGCTTCCACAGCTGTGAAGA 665  
 Db 79 uAlaLysSerProAlaGluAlaLysSer---ProGluLysAlaLysSerProValLysG1 98  
 QY 666 ACGACAGCTCTTCTGAAAAGGTGAAGAGGGGATGACACGATGACACCTCCCATAG 725  
 Db 98 uGluAlaLysSerProGluLysAlaLysSerProValLysGluGluAlaLysSerProAl 118  
 QY 726 TGACAGCGATGCGCTGACCTTGAAGAGCTTCAGAATCGCTTCGACGAGCGGGAACA 785  
 Db 118 aGluAla-----LysSerProGluLysAlaLysSerProAlaGluAl 132  
 QY 786 GGAGCGCCACTGAGAGCCCTCGAAAGGATCCAGAGTCCGCTGCGGGAAGAGCGCGGA 845  
 Db 132 aLysSerProGluLysAlaLysSerProValLysGluGluAlaLysSerProGluLysAl 152  
 QY 846 GGAGGTCCTCCGCGACACTGTGGCTCCGAG--GCAGTGCACACTGTGGCGGCTCTCT 902  
 Db 152 aLysSerProAlaGluAlaLysSerProGluLysAlaLysSerProAlaGluAlaLys 172  
 QY 903 GCCCAGTAAG-----CAGGAGCCCGAGAACGATCAGGG 935  
 Db 172 rProGluLysAlaLysSerProValLysGluGluAlaLysSerProGluLysAlaLys 192  
 QY 936 GGTGTGTCTCCAGGCTGGGAAAGATGACAGAGAGTATGTTGGAGGGAAGCGGCTCA 995  
 Db 192 rProValLysGluGluAlaLysSerProAlaGluAlaLysSerProGluLysAlaLys 212

A;Residues: 1-1110 <JAC>  
A;Cross-references: EMDL:U19361; NID:G532548; PION:AAA80106.1; PID:G632549  
C;Superfamily: neurofilament triplet H protein

Alignment Scores:  
Pred. No.: 0.00118 Length: 1110  
Score: 176.50 Matches: 113  
Percent Similarity: 35.60% Conservative: 86  
Best Local Similarity: 20.21% Mismatches: 235  
Query Match: 3.76% Indels: 125  
DB: 2 Gaps: 23

US-09-787-016A-1 (1-2610) x I51116 (1-1110)

Qy	253	GTITGGAGCAGGTATGGACACAAAAGCGCAGCCAGCAATGACGAGGCACCTAAGGCCATC	312
Db	481	ValSerAlaGluLysAspGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu	500
Qy	313	AAACCCACCAAGCAAAAGATTCGGAACAATGGGGGTTTCGAAGAGGCACCTATCGCCAAG	372
Db	501	AlaGluAlaGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu	513
Qy	373	CGAGAGGGCGAGGGGAGCGGAGCGCTGACCCACCTCGAGCGGCCACCCACACAGCAGCAG	432
Db	514	LysGluGlyGluAlaGluAlaGluGluGluGluAlaGluGluGluValGluLysGluGluAla	533
Qy	433	CTGGGCGCTTCCTCGCGCGCAGTGGAGGCGAGCCCAAGCGACTGAGCGCGTGGAGCAG	492
Db	534	GluGluAlaGluValGluGluAlaGluAlaGluGluThrGluAlaGluAlaAlaGluGlu	553
Qy	493	TTCTCTACCATTCGCGCGCGCGCGCAGAGGAGCATGCTGCTCTCTCTGGAGGATCTCT	552
Db	554	-----GluGluGluAlaGluGluGluGluGluGluGluGluGluGluGluGluGlu	562
Qy	553	GGTAGCCCACTGCTGCGCGCGCCACAGACGCGACAGCCCTCCGAGGSCACGCTGGAA	612
Db	563	AlaGluAlaGluGlyGluGluAlaGluGluAlaGluGluValGluGluAlaIleGlu	582
Qy	613	AGCGCTTCTGAGACCAAGAGCGCGCCCGCAGTCTGCTTCCACAGCTGTGAAGAACCAACA	672
Db	583	LysAlaGluAlaAla-----GluAlaLysAlaGluValGluGluGluGluGluGlu	597
Qy	673	GCCTCTTCTGAAGAGTGAAGAGGAGGTGACCCAGATGACACCTCCGATAGTCAGACGC	732
Db	598	AlaGluAlaGluGlu-----GluGluGluGluGluAlaGluGluGluGluGluVal	613
Qy	733	GATGGCTGACCTGGAAGAGCTTCAGATCGCCTT-----768	768
Db	614	GluAlaGluThrLysGluGluValGluAlaGluAlaGluValGluGluGlyGluAla	633
Qy	769	-----CGCAGGAGCGGGAACAGAGCCCACTGAGAGGCCCTGAAAGGATCCAGAGT	822
Db	634	AlaGluGluGluAlaGluGluGluGluAlaGluGluGluValThrSerLysLysAla	653
Qy	823	CGCTCTCGGAGAGCGCGCGAGGAGGTCCCGCGAGACTGTGGGCTCCGAGGGCCAGT	882
Db	654	LysThrGluGluAlaGluValGluGluGluGluAlaGluAlaAlaGluAlaGluGlu	673
Qy	883	GACACTGTGGAGGGCTCTCTGCCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTG	942
Db	674	GluGluAlaGlu-----GluGluAlaGlyGluGluAspValGlu	686
Qy	943	TCCAGGCTGGGAACATGACAGAGAGCTAAGTTGGAGGAAGCGGCTCAGGAC---	999
Db	687	AlaGluSerLysGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu	706
Qy	1000	-----ATCAAAAGATGAGGAGCTTGGAGACTTGGCGCGACCGAAGCCCTGAA	1044
Db	707	AlaGluGluGluGluValLysGluGluGluValThrLysSerAspAlaGluGluAlaGlu	726
Qy	1045	TGTGAGGGTTACGAGCCCAACGCGCTGTATGTGATTTGGCGCAGCCTCAACACAGG	1104
Db	727	AlaGluAlaGluGluGluAlaAla-----734	734

996 GGACATCAAGAATGAGGACCGCTGGAGACTTGGGCGACCGAAGCCTGAATGTGAGGGTTA 1055  
 212 rProValLysGluGluAlaLysSerProGluLysAlaLysSerPro----- 227  
 1056 CGACCCCAAGCCCTGTATTGCTATTGCTTGGCGCAGCCTCACACACAGSITTTATGATTG 1115  
 227 ----- 227  
 1116 CTGTGACCGCTGTGAAGAATGGTTTCATGCGGATTTGTGGGCAFTTCTGAGGCTCGAGG 1175  
 228 -----AlaGluAlaLysSe 232  
 1176 GAGGCTTTGGAAAGAAATGGGAGAGACTATATCTGCCCAAACCTGCACCAATCTCTCAAGT 1235  
 232 rProValLysGlu-----GluAl 238  
 1236 GAGAGATGAGACTCATTTAGAAACCGCAGATCAGCAGGAAGCTAAATGAGAGACCTGGAGA 1295  
 238 alySerProGluLysAlaLysSerProGluLysGluGluAlaLys---SerProAlaGl 257  
 1296 TGCTGATGGACCGATTGTACAAATATAGAACATAGACGAGAGACTCTAGCGAAGACCA 1355  
 257 uAlaLysSer-----ProGluLysAlaLysSerProGl 268  
 1356 AGGGATAAAGGGTAGAATTGAGAAAGCTGCAAAATCAAGTGGCAAGCAAGAACTCAAGAT 1415  
 268 uLysAlaLysSerProValGluValLysSerProAlaGluAlaLysSerProGluLysAl 288  
 1416 CTTCCAGCTGTG---ATAGAGCGCTGTGCTCAAAATGTATTGGCCCGGGTGTG 1472  
 288 alySerProValLysGluGluAlaLysSerProGluLysAlaLysSerProValLysGl 308  
 1473 TCAGTGGCGCAGCCGACTCGGTGTACTGCTAGTAATGACTGTATCTCAAACACGCGC 1532  
 308 uGluAlaLysSerProGluLys-----AlaLysSerProValLysGluGluAlaLys 325  
 1533 AGCGACAATCAAGTTTCTTAAGCTCAGGTAAGAACACAGAACCCAGCCCT---AAAGAAA 1589  
 325 sSerProGluLysAlaLysSerProValLysGluGluAlaLysSerProGluLysAlaLys 345  
 1590 GATGAGATGACAGCAGAGAGCCAGCTCTCCGAATCGGTGCTCAGCGAGCTATTAA 1649  
 345 sSerProValLysGluGluAlaLysSerProGluLysAlaLysSerProValLysGluGl 365  
 1650 AATCTCTTCTGTGCACAAAGAGACCGACTCCAGAAAAAAGAGACACACAGTGAAGAAGGC 1709  
 365 uAlaLysSerProGluLysAlaLysSerProGluLysAlaLysSerProValLysGluGl 385  
 1710 AGTGTGTGTCCTGCGCGGAGTGAAGCACTTCGGGAAGGAGCAGCTTGTGAGAGAGCAAC 1769  
 385 uAlaLysSerProGluLysAlaLysSerProValLysGlu-----GluAlaLysSe 402  
 1770 GCGTGTGCGCGGCGGATGCACAAATCAATGAGTAAAGCCAGAAAGACTGCTGCTCC 1829  
 402 rProGlu-----LysAlaLysSerProValLysGluGluAlaLys 415  
 1830 CTGCGCG 1836  
 415 sSerPro 417  
 RESULT 8  
 I51116  
 NF-180 - sea lamprey  
 C:Species: Petromyzon marinus (sea lamprey)  
 C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 21-Jul-2000  
 C:Accession: I51116  
 R:Jacobs, A.J.; Kamholz, J.; Selzer, M.E.  
 Brain Res. Mol. Brain Res. 29, 43-52, 1995  
 A:Title: The single lamprey neurofilament subunit (NF-180) lacks multiphosphorylation re  
 A:Reference number: I51116; MUID:95287814; PMID:7770000  
 A:Accession: I51116  
 A:Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Volecule type: mRNA

QY 1105 TTTATGTTGCTGTGACCGCTGTGAGGATGTTTTCATGCGGATGTTGTTGGCAATTTCT 1164  
Db : : : : :  
QY 735 : : : : :  
Db : : : : :  
QY 1165 GAGGCTCGAGGAGGCTTTTGAAGAAGGATGGAAGAGACTATATCTGCCAAACTGCACC 1224  
Db : : : : :  
QY 743 GluAlaLysAspGluAlaGluGluGluAlaGluGlu : : : : : 755  
QY 1225 ATTCTGCAAGTGCAGGTGAGTCTCATTGAGAAACCGCATCAGCAGGAAGCTAAATGG 1284  
Db : : : : :  
QY 756 : : : : :  
Db : : : : :  
QY 1285 AGACCTGGAGTGTGATGTCACCGCTTCCAGCCTGTGATAGAG : : : : : 1344  
Db : : : : :  
QY 773 AlaSerAspGluLysProGluGlu : : : : : 785  
QY 1345 AGCGAA : : : : : 1392  
Db : : : : :  
QY 786 SerGluAlaProValAlaProGluAlaLysAlaProGluProLysAla : : : : : 804  
QY 1393 AGTGGCAAGAAAGAACTCAAGATCTCCAGCCTGTGATAGAG : : : : : 1443  
Db : : : : :  
QY 805 LysLysLysAlaProAlaLysValGluSerProThrSerGluProGluAspGluProLys 824  
QY 1444 GCCTCAAAATGTATGGCCCGGCTGTGTCAGTGGCGCAGCCGACTCGGTGTACTGC 1503  
Db : : : : :  
QY 825 AlaGluValGluLysLysGlyLysAlaGluAlaProLysProLysAla : : : : : 841  
QY 1504 AGTAATGACTGTATCTCAACACCGCGCAGCAGCAATGAGTTTCTAAGCTCAGGTAA 1563  
Db : : : : :  
QY 842 : : : : : 855  
QY 1564 GAACAGAGCAAAAG : : : : : 1617  
Db : : : : :  
QY 856 GluGluLysProGluGluSerProThrGluGlu : : : : : 872  
QY 1618 CTTCCGAATGCGGTGTGAGCGAGGTATTAATCTCTTCTGTGCAAGAGACCACT 1677  
Db : : : : :  
QY 873 AlaAlaLys : : : : : 886  
QY 1678 CCAGAAAGAAAGACACAGTGAAGAGGAGGAGTGTGCTGCGCGAGTGA : : : : : 1734  
Db : : : : :  
QY 887 Pro : : : : : 905  
QY 1735 GCATCGGAGGAAGCAGCTGTGAGAGCAGCAGCCGCTGCTGGGCGAGCATCAAT 1794  
Db : : : : :  
QY 906 AlaProAlaAlaGluGluGluAspGluLysGluAspGluGluGluGluGlu 925  
QY 1795 TACAATGCATGAAGCCAGAAAGACT : : : : : 1839  
Db : : : : :  
QY 926 ValGluGluValLysProGluAspAlaLysProValLysSerLysProAlaProAla 944

RESULT 9  
B42680  
Nucleolus-cytoplasm shuttle phosphoprotein - rat  
N;Alternate names: Nopp140 protein B  
C;Species: Rattus norvegicus (Norway rat)  
C;date: 30-Sep-1993 #sequence, revision 30-Sep-1993 #text change 17-Nov-2000  
C;Accession: B42680; A42680; S27889; S30510; S30511  
R;Meier, U.T.; Blobel, G.  
Cell 70, 127-138, 1992  
A;Title: Nopp140 shuttles on tracks between nucleolus and cytoplasm.  
A;Reference number: A42680; MUID:92323542; PMID:1623516  
A;Accession: B42680  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-734 <ME1>  
A;Cross-references: GB:M94288  
A;Experimental source: clone pTM6  
A;Accession: A42680  
A;Status: preliminary  
A;Molecule type: mRNA

A;Residues: 32-180,'Q',181-734 <ME3>  
A;Cross-references: GB:M94287; NID:g205749; PIDN:AAA41718.1; PID:g205750  
R;Meier, U.; Blobel, G.  
submitted to the EMBL Data Library, May 1992  
A;Description: Nopp140 shuttles on tracks between nucleolus and cytoplasm.  
A;Reference number: S27889  
A;Accession: S27890  
A;Molecule type: mRNA  
A;Residues: 32-734 <ME2>  
A;Cross-references: EMBL:M94288; NID:g205751; PIDN:AAA41719.1; PID:g205752  
A;Accession: S27889  
A;Molecule type: mRNA  
A;Residues: 32-180,'Q',181-734 <ME12>  
A;Cross-references: EMBL:M94287; NID:g205749; PIDN:AAA41718.1; PID:g205750  
C;Superfamily: nucleolus-cytoplasm shuttle phosphoprotein  
C;Keywords: phosphoprotein

Alignment Scores:  
Pred. NO.: 0.00163 Length: 734  
Score: 174.00 Matches: 132  
Percent Similarity: 35.42% Conservative: 77  
Best Local Similarity: 22.37% Mismatches: 250  
Query Match: 3.71% Indels: 131  
DB: 25 Gaps: 25

US-09-787-016A-1 (1-2610) x B42680 (1-734)

QY 172 TCTGCTTGACCAAGAGGTTTCGTCCTCCAGGCTTTTGGTGTATTTAGATTTCAGGG 231  
Db : : : : :  
QY 86 SerPheTrpLeuLysSerThrLysAlaPro : : : : : 103  
QY 232 AAAAGTGTCAAGCTTTCAGTGTGGAGCAGGTATGAGCAGCAAGGCGACCCGCAAT 291  
Db : : : : :  
QY 104 ProValAlaLysAlaLysLysGluThrSerSerSerSerSerSerSerSerSerSer 123  
QY 292 GAGGAGGACCTTAAGGCCATCAAAACCCAGCAGCAAGAGTTCAGGAAACATCGGGGTTT 351  
Db : : : : :  
QY 124 GluGluGluAspLysAlaGlnValProThrGlnLys : : : : : 135  
QY 352 CGAAGACCACTATGTCGCAAGCA : : : : : 399  
Db : : : : :  
QY 136 ---AlaAlaAlaProAlaLysArgAlaSerLeuProGlnHisAlaGlyLysAlaAla 154  
QY 400 GACCCACTGAGCGCGCCACCCACAGCAGCAGCTGGGCTGTCCCTGCGGCGCAGTGGG 459  
Db : : : : :  
QY 155 LysAlaSerGluSerSerSerSerSerGlu : : : : : 170  
QY 460 AGGCAGCCCAAGCAGCAGCTGAGCGCTGGAGCAGTCTCTGACCATTTGCGGCGCGCGGC 519  
Db : : : : :  
QY 171 GluLysAspLysLysLysProValGlnLysAlaValLysProGlnAlaLysAlaVal 190  
QY 520 AGGAGGAGCATCTCTCTCTCTGAGGATTCGTGTGAGCCACGCTCTCTGCGGCGCACA 579  
Db : : : : :  
QY 191 Arg : : : : : 208  
QY 580 GACGCGGAGCAGCCTCCGAGGCGCAGCTTGGAAGCGCTTCTGAGACCAAGAGCGCCCC 639  
Db : : : : :  
QY 209 GluAspGluAlaProGlnThrGlnLysProLysAlaAlaAlaAlaAlaAlaAlaPro 228  
QY 640 CAGTCTGCTTCCACAGCTGTGAAGAGCAGCAGCCTCTCTGAAAAGTG : : : : : 690  
Db : : : : :  
QY 229 ThrLysAlaGlnThrLysAlaProAlaLysProGlyProProAlaLysAlaGlnProLys 248  
QY 691 : : : : : 738  
Db : : : : :  
QY 249 AlaAlaAsnGlyLysAlaGlySerSerSerSerSerSerSerSerSerSerSerSer 268  
QY 739 CTGAGCTGGAAGAGCTTCAGAACTCGCTTCGAGGAAGCGGGAACAGAGGCCCACTGAG 798  
Db : : : : :  
QY 269 : : : : : 276  
QY 799 AGGCCCTGAAAGGATCCAGAGTCGCTCGGAGAGCGCGCGGAGGAGGTCCCGCC 858



277	AlaProLeuLys	-----LysThrAlaProLys	285
859	GAGACTGTGGGCTCCGAGGCCAGTGACACTGTGGAGGGCGTCTCGCCCAAGTAAGCAGGAG	918	
286	LysGlnValValAlaLysAlaProValLysValThrAlaAlaProThrGlnLysSerSer	305	
919	CCCGAGAACCATCAGGGGTGTGTCCCGCTGGGAAGATCACAGAGAGACTAGTTG	978	
306	SerSerGluAsp	-----SerSerSerGluGluGluGluGluGlnLysLysPro	321
979	GAGGGAAGCGGCTCAGGACATCAAGATGAGGAGCCTGGAGACTTGGGCCACCGAAG	1038	
322	MetLysLysLysAlaGlyProTyrSerSerValProProSerValSerLeuSerLys	341	
1039	CCTGAATGTGAGGGTTACGACCCCAAGCCCTGTATTGCAATTGGCCGACAGCTCACAA	1098	
342	LysSerValGlyAlaGlnSerProLysLysAla	---AlaAlaGlnThrGlnProAlaAsp	360
1099	AACAGGTTTATGTTCTGTGACCGCTGTGAAGAATGGTTTCATGGCGATTGTGTGGGC	1158	
361	Ser	-----SerAlaAspSerSerGluGlu	368
1159	ATTCTGAGGCTCGAGGAGCGCTTTTGAAAGAAATGGGGAAGACTATATCTGCCCAAC	1218	
369	---SerAspSerSerGlu	-----GluGluLysLysThrProAla	381
1219	TGCACCATTTCCGAGTGCAGGATGAGACTCATTCCGAACCGCAGATCAGCAGGAAGCT	1278	
382	LysThrValVal	-----SerLysThrProAlaLysProAlaProValLysLysLys	398
1279	AAATGGAGACTCGAGATGCTGTGCGCAGCATTTGTACAAGTATAGGAACATAGAGCAG	1338	
399	AlaGluSerSerSerAspSerSerAspSerAsp	-----	409
1339	AAGTCTGAGCAAGCCAGGATTAAGGGTAGAATTGAGAAGCTGCMAATCCAACTGGC	1398	
410	---SerSerGluAsp	-----GluAlaProAlaLysProValSer	421
1399	AAGAAGAAACTCAAGATCTCCAGCCT	---GTGATAGAGCGCGCTGTGTCTCAAAATGT	1455
422	AlaThrLysSerProLeuSerLysProAlaValThrProLysProProAlaAlaLysAla	441	
1456	ATTGGCCCCGGTCTGTCTGCTGCGCAGCCCGACTCGGTGTACTCCAGTATGACTGT	1515	
442	-----ValaThrProLysGlnProAlaGlySerGlyGln	--L	454
1516	ATCCTCAACACGCGCAGCGACAAAT	-----GAAGTTCTAAGCTCAGGTAAA	1563
454	ysProGlnSerArgLysAlaAspSerSerSerSerSerGluGluGluSer	---SerSerSerGlu	473
1564	GAACAGAACGCCAAG	-----CCTAAAGAAAAGATGAAGATGAAGCCAGAG	1608
474	GluGluAlaThrLysLysSerValThrThrProLysAlaArgValThrAlaLys	---Ala	492
1609	AAGCCCACTTCCG	---AAATGGGTGCTCAGCAGGTATTAAATCTCTTCTGTGCGC	1665
493	AlaProSerLeuProAlaLysGlnAlaProArgAlaGlyLysAspSerSerSerAspSer	512	
1666	AAGAGACCACTCCAGAAAAAAGAGACACCACTGAAGAGGCGAGTGGTGGTCCCTCGC	1725	
513	GluSerSerSerSerGluGluGluLysLysLysThrProProLysPro	-----ProAla	529
1726	CGGAGTGAGCACTCGG	-----AAGGAAGCAGCTTGT	1758
530	LysLysLysAlaAlaGlyAlaAlaValProLysProThrProValLysLysAlaAlaAla	549	
1759	GAGAGCAGCAGCCGCTCGTGGCGGCAGGCATCACAACT	---TACAATGCAAGTAAAGCCA	1812
550	GluSerSerSerSerSerSerSerSerSerGluAspSerSerGluGluGluLysLysPro	569	
1813	GAAGAAGCTGTGCTCCCTCGCGG	1836	
570	LysSerLysAlaThrProLysPro	577	

RESULT 10

KRS001110  
 T21430  
 hypothetical protein F26H11.3b - *Caenorhabditis elegans*  
 C:Species: *Caenorhabditis elegans*  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 02-Sep-2000  
 C:Accession: T21430  
 R:Barlow, K.  
 submitted to the EMBL Data Library, November 1996  
 A:Reference number: Z19421  
 A:Accession: T21430  
 A:Status: preliminary; translated from GB/EMBL/DBDJ  
 A:Molecule type: DNA  
 A:Residues: 1-510 <WIL>  
 A:Cross-references: EMBL:Z81515; PIDN:CA804195.1; GSPDB:GN00020; CBSP:F26H11.3b  
 A:Experimental source: clone F26H11  
 C:Genetics:  
 A:Gene: CBSP:F26H11.3b  
 A:Map position: 2  
 A:Introns: 116/1; 147/3; 288/2; 348/2; 392/1  
 C:Superfamily: bromodomain homology  
 P:371-426/Domain: bromodomain homology <ERO>

[illegible]

Pred. No.:	0.00332	Length:	990
Score:	169.50	Matches:	120
Percent Similarity:	34.08%	Conservative:	63
Best Local Similarity:	22.35%	Mismatches:	216
Query Match:	3.61%	Indels:	138
DB:	2	Gaps:	22

US-09-787-016A-1 (1-2610) x I51618 (1-990)

Qy	271	GACAAAGGCGACCCGAGCGAATGAGGAGGCACCTAAGGCATCAAAACCACCAAGAG	330
Db	390	AssSerAspSerSerAsnGluThr-----ThrThryLysProAlaAlaLys---	406
Qy	331	TTCCAGGAAACATGGGGTTTTCGAGGACCACTATCCCAAGCGAGAGCGCGCAGGGGAC	390
Db	407	-----ThrThrProAlaLys-SerAlaAlaThrProTh	417
Qy	391	GGGAGGCTGACCCACT---GGAGCGCCACCCACAGCAGCAGCAGCTGGGCCGTCCCTG	447
Db	417	rSerLysThr-ProThrAsnGlyLysAlaThrProThrSer-----LysThrProAla	434
Qy	448	CGGCGCAGTGGGAGGAGCCCAAGCAGCAGCGCGGTGGAGCAGTTCCTCGACCATTCGG	507
Db	434	la-LysProGly---ThrProLys-----ThrSerThrAla	444
Qy	508	CGGCGCGCGCGCAGGAGGAGCAGTCCTGTCCTCCCTGGAGGATTCGTGTGAGCCCAAGTCC	567
Db	445	LysLysAspSerSerSerAspSerSerSerAspSerSerSerAspGluGluThrThrThr	464
Qy	568	TGCCCGCC-----ACAGACCGCGACAGCAGCTCCGAGGCGAGCTGGGAAAGCGCT	618
Db	465	LysProAlaAlaLysThrThrProAlaLysSerAlaAlaAlaThrProThrSerLysThrPro	484
Qy	619	TCTGAGACCAAGAGCGGCCCTGCTGCTTCCACAGCTCTGTGAAGGAACGACCGCTCTCT	678
Db	485	ThrAsnSerLysAlaThrProThrSerLysLysThrProAlaLysProGlyThrProLys	504
Qy	679	TCTGAAAGGTGAAGAGCGGGGTGACCAAGTACACACTCCGATAGTGCAGCAGGTGGC	738
Db	505	ThrSerAlaAlaLysLysAspSerSerSerSerSerSerSerSerSerSerSerSerAsp	523
Qy	739	CTGACCTTGAAGAGAGCTTCAGATGCGCTTCGAGGAAGCGGAACAGGAGCCGCTGAG	798
Db	524	-----GluLysLysThrProAlaLysArgAlaLysThrThrProAlaLys	539
Qy	799	AGGCCCTGAAAGGGATCCAGATCGCGCTCGGAGAGAGCCCGGAGGAGGCTCCCGCC	858
Db	540	ProAlaAlaLysThrThrProAlaLysProAlaAlaLysThrThrProAlaLysProAla	559
Qy	859	GAGACTGTGGCTCCGAGGCCAGTGACACTGTGTGAGGGCGCTCTGCCAGCTAAGCAGAG	918
Db	560	-----AlaLysSerThr-----ProGlyLysGlnVal	568
Qy	919	CCGAGAACGATCAGGGGGTTGTGTCACGCTGGGAAAGATGACAGAGAGTAAAGTTG	978
Db	569	ProThrLysLysGluSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer	588
Qy	979	GAGGGAAGGGCGCTCAGGCATCAAA-----GATGAGGAGCCTGGA	1020
Db	589	LysSerSerAlaLysProAlaValLysThrThrProGlyLysAlaThrSerLysProVal	608
Qy	1021	GACTTGGGCGACGAAGCCTGAATGTGAGGGTTACGCCCAAGCCCTGTATTGCAAT	1080
Db	609	ValAlaSerLysProValProAlaLysLysAlaSerSerSerSerSerSerSerSer	624
Qy	1081	TGCCGCCAGCTCAACAACACAGGTTATGATTTGCTGTCACGCTGTGAAGATGCTTT	1140
Db	625	-----SerAspSerSerGluGluLysThr	632
Qy	1141	CATGGCGATTGTGGGCATTTCTGAGGCTGACGAGGAGGCTTTTTCGAAAGGATGGGAA	1200
Db	633	ThrLysThrThrLysProLysThrLys-----	641

QY 1201 GACTATATCTGCCAACTGCACCACTTCTG-----CAAGTGCAGGATGAGACTCAT 1251  
 Db 642 -----LeuSerProAlaValLysThrLeuProProLysLysAlaGluSerSerAsp 659  
 QY 1252 TCAGAAACGGCAGATCAGCAGCAAGCTAAATGAGACCTGAGAGCTGTGATGGCAGCAT 1311  
 Db 660 SerSerAspSerAspSerGluLysLysThrLysProAlaLysProAlaLysSer 679  
 QY 1312 TGTCAAGTATAGGAACA----- 1329  
 Db 680 AlaThrProValAsnThrLysAlaProAlaGlnAsnLysAlaSerLysAlaSerCysSer 699  
 QY 1330 ATAGAGCAGAAGCTTAGCGAAGACCAAGGGATAAGGGTAGAATTGAGAAAGCTGCAAT 1389  
 Db 700 AspSerAspSerSerSerGluGluGluGly-----LysSerLysGln 713  
 QY 1390 CCAAGTGGCAAGCAAGAACTCAAGATCTCCAGCTGTGTAGAGGCGCTGGTGCCTCA 1449  
 Db 714 ProThrGlyLys-----SerProAlaAlaLysAlaThrAlaProPro 727  
 QY 1450 AAATGTATTGGCCCGGTGTGTCACTGCGCAGCCGCACTCGGTGTACTGCAAT 1509  
 Db 728 Lys-----LysAsnProValAlaValAsnLysAspLysProSerSerSerSerAsp 746  
 QY 1510 GACTGTATCTCAACACGCGCAGCATGATGAATTTCTAAGCTCAGGTAAAGACAG 1569  
 Db 747 -----SerSerGlyAspAspGlu 752  
 QY 1570 AAGCCAAAGCTTAAGAA-----AAGATCAAGATGAAGCCAGAG 1608  
 Db 753 LysGlnLysProLysGlnAlaAlaAlaLysAspValLysGlnGlyAlaLysAlaAla 772  
 QY 1609 AAGCCAGTCTCCGAATGGGTGCTCAGCAGGATTAATAATCTCTGTGTGCACAG 1668  
 Db 773 LysProThr-----ProLysAlaAlaLysSerSerSerSerGluAspSerSerAspGlu 791  
 QY 1669 AGACAGCTCCAGAAAGAAAGACACCACTG-----AAGAAAGGAGTGGTGCCT 1722  
 Db 792 AspValSerLysAlaLysLysThrAsnThrAlaValSerLysSerProValThrPro 811  
 QY 1723 GCGCGAGTGAAGCTCGGAGAGCAGCACTTGTGAGCAGC 1767  
 Db 812 LysAlaValProAlaAlaLysLysGluSerSerSerGluSerSer 826  
 RESULT 12  
 JW0057  
 C:Species: Homo sapiens (man)  
 C:Date: 13-Jun-1998 #sequence\_revision 10-Jul-1998 #text\_change 08-Oct-1999  
 C:Accession: JW0057  
 R:Sato, N.; Kokame, K.; Shimokado, K.; Kato, H.; Miyata, T.  
 J. Biochem. 123, 1119-1126, 1998  
 A:Title: Changes of gene expression by lysophosphatidylcholine in vascular endothelial cells  
 A:Reference number: JW0057; MUID:98269042; PMID:9604001  
 A:Accession: JW0057  
 A:Molecule type: mRNA  
 A:Residues: 1-1684 <SAT>  
 A:Cross-references: DBJ:AB003476; NID:g2081606; PIDN:BAAL19927.1; PID:d1020716; PID:g2081606  
 C:Comment: This protein regulates cell growth.  
 F:433-439/Region: nuclear location signal  
 F:522-527/Region: nuclear location signal  
 F:591-596/Region: nuclear location signal  
 F:671-676/Region: nuclear location signal  
 Alignment Scores:  
 Pred. No.: 0.00355 Length: 1684  
 Score: 169.50 Matches: 139  
 Percent Similarity: 30.58% Conservative: 76  
 Best Local Similarity: 19.77% Mismatches: 221  
 Query Match: 3.61% Indels: 267  
 DB: 2 Gaps: 30  
 US-09-787-016a-1 (1-2610) x JW0057 (1-1684)

QY 226 TCAGGAAAGTGTCCAGCTTTCAGTGTGGACAGATATGGACAGCAAGGCGACCCG 285  
 Db 160 SerGlyGlnAlaValGluGlu-----CysLysGluGluGluGlu 173  
 QY 286 AGCAATAGGAGGCACTT-----AAGCCATCAAAACCCACAGC- 324  
 Db 174 LysGlnGluLysGluProSerLysSerAlaGluSerProThrSerProValThrSerGlu 193  
 QY 325 -----AAGAGTTCAGGAAACATGG-----GGTTTTCGAGGACCACT 363  
 Db 194 ThrGlySerThrPheLysLysPheThrGlnGlyTrpAlaGlyTrpArgLysLysThr 213  
 QY 363 ----- 363  
 Db 214 SerPheAspLysProLysGluValGluAlaSerGluLysLysLysGluGlnGlu 233  
 QY 364 -----ATCGCAAGCGAGAGGCGCA 384  
 Db 234 ProGluLysValAspThrGluGluAspGlyLysAlaGluValAlaSerGluLysLysThr 253  
 QY 385 GGGACGCGGAGGCTGACCCACTGGAGCGGCA----- 417  
 Db 254 AlaSerGluGlnAlaHisProGlnGluProAlaGluSerAlaHisGluProArgLysSer 273  
 QY 418 -----CCCCACAGCAGCAGCTGGGCTGTCCCTCGCGGCGC 453  
 Db 274 AlaGluTyrgLysValGluValGluLeuProSerGluGluGlnVal----- 287  
 QY 454 AGTGGGAGGCGAG-----CCCAAGCGCACTGAGCGGTGGAG 489  
 Db 288 SerGlySerGlnGlyProSerGluGluLysProAlaProLeuAlaThrGluValPheAsp 307  
 QY 490 CAGTTCCTGACCATTCGCGCGCGCGCGAGGAGCATGCTGCTCCCTCGGAGAT 549  
 Db 308 GlnLysIleGluValHisGlnGluValValAlaGluValHisValSerThrValGlu 327  
 QY 550 TCTGTGAGCCGACCTCTCTCCGCGCACAGAGCC-----GAGACAGCCTCCGAGGCGCAGC 506  
 Db 328 -----GluArgThrGluGluGlnLysThrGluValGluGluThrAlaGlySerValPro 345  
 QY 607 GTGGAAAGCGCTTCAGACACAGAGCGCCCGCTGCTGCTCCACAGCTGTGAGGAA 666  
 Db 346 AlaGluGluLeuValGluMetAspAlaGluProGlnGluAlaGluProAla-----LysGlu 364  
 QY 667 CGACCAAGCTCTTCTGAAAGGTGAAAGGAGGGATGACACAGATGACACCTCCGATAGT 726  
 Db 365 LeuValLysLeuLysGluThrCysValSerGlyGluAspProThrGlnGlyAlaAspLeu 384  
 QY 727 GACAGCATGGCTGACCTTGAAGAG----- 753  
 Db 385 SerProAspGluLysValLeuSerLysProGluGlyValValSerGluValGluMet 404  
 QY 754 CTTCAAGATCGCTTCGAGGAAAGCGGAAACAGAGAGCC-----ACTGAG 798  
 Db 405 LeuSerSerGlnGluArgMetLysValGlnGlySerProLeuLysLysLysLysPheThrSer 424  
 QY 799 AGGCCCTTGAAGGATTCAGATGCTCGCGGAGAGAGCGCGGAGGAGGT----- 852  
 Db 425 ThrGlyLeuLysLysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyAspGlu 444  
 QY 853 -----CCCCCGAGACTGTGGCTCGAGGCGCAGTGCACACT 888  
 Db 445 GluSerGlyGluHisThrGlnValProAlaAspSerProAspSerGlnGluGlnLys 464  
 QY 889 GTGGAGGCGCTCTCCCGCAGTAAGCAGGAGCCGAG----- 924  
 Db 465 GlyGluSerSerAlaSerSerProGluGluProGluGluLeuThrCysLeuGluLysGly 484  
 QY 925 -----AAGCATCAGCGGCTTGTGTCAGGCTGGGAAAGATGACAGAGAGT 972  
 Db 485 LeuAlaGluValGlnGlnAspGlyGluAlaGluGluGlyAlaThrSerAspGlyGluLys 504

QY 973 AAGTTGGAGGGA----- 984  
Db 505 LysArgGluGlyValThrProTrpAlaSerPheLysLysMetValThrProLysLysArg 524  
QY 984 ----- 984  
Db 525 ValAla-gArgProSerGluSerAspLysGluAspGluLeuAspLysValLysSerAlaThr 544  
QY 985 -----AAGCGCGCTCAGGACATCAAGATGAGAGCGCTGAGACTTGGGC 1029  
Db 545 LeuSerSerThrGluSerThrAlaSerGluMetGlnGluMetLysGlySerValGlu 564  
QY 1030 CGACCGAGCTGAATGTCAGGGTTACGACCCCAAGCCCTGTATTGCTATTTGCCGCCAG 1089  
Db 565 GluProLysProGlu-----Glu 570  
QY 1090 CCTCACAAACAGAGGTTTATGATTCTGTGACCGCTGGAAGAATCGTTTCATGCGCAT 1149  
Db 571 ProLysArgLys-----ValAspThrSerValSerTrpGluAlaLeuIle 585  
QY 1150 TGTGTGGGCAATTTCTAGGCTCGAGGAGGCTTTTGGAAAGNATGGGAAGACTATATC 1209  
Db 586 CysValGlySerSerLysLysArgAla-----ArgArgGlySerSer----- 599  
QY 1210 TGCCCAAACTGCACCATTTCTGCAAGTGCAGGATGAGACTCAATTCAGAAACGGCA----- 1263  
Db 600 -----SerAspGluGluGlyProLysAlaMetGly 610  
QY 1264 ---GATCAGCAGGAAGCTAAATGAGACCTGGAGATGCTGAT---GGCACGATTGTACA 1317  
Db 611 GlyAspHisGlnLysAlaAspGluAlaGlyLysAspLysGluThrGlyThrAspGlyIle 630  
QY 1318 AGTATAGGACAAATAGCAGNAGTCTAGCGAAGACCAAGCGATAAAGGTAGATTGAG 1377  
Db 631 LeuAlaGlySer-----GlnGluHisAspProGlyGlnGly---SerSerSerProGlu 647  
QY 1378 AAAGCTGCAATCAAGTGGCAAGAAACTCAAGATCTTCCAGCTGTGTAGAGCGG 1437  
Db 648 GlnAlaGlySerProThr-----GlnGly 655  
QY 1438 CTGTGTGCTCAAAATGTTATTTGGCCCGGTGCTGTACGTGGCGCAGCCGACTCGTG 1497  
Db 656 GluGlyValSer----- 659  
QY 1498 TACTGCAGTAATGACTGTATCTCCAAACAGCCCGCAGCAATGAAGTTTCTAAGTCA 1557  
Db 660 -----ThrTrpGluSerPheLysValGluValThr 669  
QY 1558 GGTAAAGACAGAACCCCTAAAGAAAGATGAGATGAGCCAGAGAGCCCGAGT 1617  
Db 670 -----ProArgLysLysSerLysSerLysGluGluLysSer 682  
QY 1618 CTTCCGAAATGCGGTGCTCAGCGAGGTATTAAATCTCTTGTGCAACAGAGACCAAGCT 1677  
Db 683 GluAspSerIle---AlaGlySerGlyValGluHisSerThrProAspThrGluProGly 701  
QY 1678 CCAGAAATAAAGAGACCAAGTGAAGAGGACATGTTGTCCTGCGCGGAGTGAAGCA 1737  
Db 702 LysGluGluSerTrpValSerIleLysLysPheIleProGlyArgLysLysArgPro 721  
QY 1738 CTCGGGAG-----GAAGCAGCTGTGAGAGCAGCAGCGCTGTGTCGGAGGAGATCAC 1791  
Db 722 AspGlyLysGlnGluGlnAlaProValGluAspAlaGlyProThrGlyAlaAsnGluAsp 741  
QY 1792 AAT-----TACATGCAAGTAAAGCCAGAAAG 1818  
Db 742 AspSerAspValProAlaValValProLeuSerGluTyrAspAlaValGluArgGluLys 761  
QY 1819 ACTGCTGCT 1827  
Db 762 MetGluAla 764

RESULT 13

S68191 - human  
C:Species: Homo sapiens (man)  
C:Date: 15-Feb-1997 #sequence\_revision 13-Mar-1997 #text\_change 11-Jan-2000  
R:Accession: S68191  
R:Task: N.I.; Eyre, H.J.; O'Brien, R.O.; Sutherland, G.R.; Denborough, M.A.; Foster, P.  
Bur. J. Biochem. 233, 258-265, 1995  
A:Title: Molecular cloning of the cDNA encoding human skeletal muscle triadin and its lo  
A:Reference number: S68191; MUID:96061957; PMID:7588753  
A:Accession: S68191  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-729 <TAS>  
A:Cross-references: EMBL:U18985; NID:9882222; PIDN:AAA75315.1; PID:9882223  
C:Genetics:  
A:Map position: 6q22-6q23  
C:Superfamily: histone H1  
  
Alignment Scores:  
Pred. No.: 0.00345 Length: 729  
Score: 169.00 Matches: 101  
Percent Similarity: 35.11% Conservative: 83  
Best Local Similarity: 19.27% Mismatches: 168  
Query Match: 3.60% Indels: 172  
DB: 22 Gaps: 22  
  
US-09-787-016A-1 (1-2610) x S68191 (1-729)  
QY 532 CTGTCTCTCCCTGAGGATTTCTGTGAGCCACGCTCTGCGCCGACACGAGCCGAGACA 591  
Db 293 ProLeuProThrGluGlnAlaSerArgProThr-----ProAlaSerProAlaLeuGlu 310  
QY 592 GCTCTCCGAGGCGCTGGAAAGCGCT-----TCGAGACC----- 527  
Db 311 GluLysGluGlyGluLysLysLysAlaGluLysLysValThrSerGluThrLysLysLys 330  
QY 628 -----AGAGCGGCCCGCCAGTCTGCTTCCACAGCTGTAAGGAA 666  
Db 331 GluLysGluAspIleLysLysLysSerGluLysGluThrAlaIleaspValGluLysLys 350  
QY 667 GCACAGACCTCTTCTGAAAGAGTGAAGGAGG----- 699  
Db 351 GluProGlyLysAlaSerGluThrLysGlnGlyThrValLysIleAlaAlaGlnAlaAla 370  
QY 700 -----GATGACACCATGACACCTCGGATAGTACAGCGATGCGCTTGAA 750  
Db 371 AlaLysLysAspGluLysLysGluAspSerLysLysThrLysLysProAlaGluValGlu 390  
QY 751 GAGCTTCAGAAATCGCTTCGCGAGAGCGGAGAACAG-----GAGCCCATGAGAGGCC 804  
Db 391 GlnProLysGlyLysLysGlnGluLysLysGluLysHisValGluProAlaLysSerPro 410  
QY 805 CTGAAA-----GGGATCCAGATCCCTCGGAGAGAGCGCGGAGGAGGTCGCCGCC 858  
Db 411 LysLysGluHisSerValProSerAspLysGlnValLysAlaLysThrGluArgAlaLys 430  
QY 859 GAGACTGTGGCTCCGAGGCGAGTACACTGTGAGAGCGGCTCTGCGCCAGTAAGCAGGAG 918  
Db 431 GluGluIleGlyAlaValSerSer-----LysLysAlaValProGlyLysLysGlu 447  
QY 919 CCGAGACGATCAGGGGTTGTGCCAGCGCTGGGAAAGATGACAGAGAGTAAGTTG 978  
Db 448 GluLysThrThrLysThrValGluGlnGluIleArgLysGluLys----- 462  
QY 979 GAGGGAAGCGGCTCAGGACATCAAGATGAGAG----- 1014  
Db 463 SerGlyLysThrSerSerIleLysLysAspLysGluProIleLysGlyLysGluLys 482  
QY 1015 ---CCTGGAGACTTGGCGGACCGAAGCTGAAATGTGAGGTTTACGACCCCAAGCCCTG 1071  
Db 483 ValProAlaSerLeuLysGluLysGluProGluThrLys----- 495  
QY 1072 TATTGCATTTGGCGCCAGCTCACAACAGGTTTATGATTGCTGTGACCGCTGTGAA 1131





```
QY 598 -----GAGGGCAGCGTGGAAAGCGCTTCTGAGACGAGAGCGGCC 639
Db 949 SerProLysArgLysMetGluGlyValGluGlyLysLeuGluArgThr- 965
QY 640 CAGTCTGCTTCCAGAGCTGTGAAGAGACGACGAGCTCTTCTGAAGAGGTGAAGAGGG 699
Db 966 GluLysAspLysLeuAlaSerThrThrProAla-----LysLeuLysLeuAsn 983
QY 700 GATGACACGATGACACCTCCGATAGTACACGATGCGCTGACCTTGAAGAGCTTCAG 759
Db 984 ArgGlu-----ThrGlyLysLysLeuGly 991
QY 760 AATCGCTTCGAGAGCGGGAACAGGAGCCACTGAGAGCCCTGAAAGGATCCAG 819
Db 992 AsnAlaGluAsnAlaSerThrThrLysGluProSerGluLys---LeuGluSerThrSer 1010
QY 820 AGTCGCTG-----CGAAGAAGCGCGGAGGAG 849
Db 1011 SerLysLysLeuGluLysValLysGlyLysAlaLysArgLysValAlaGlySerGlu 1030
QY 850 GGTCCGCGCGAGACTGTG----- 867
Db 1031 GlySerSerThrLeuValAspTyrThrSerThrSerThrGlyGlySerProVal 1050
QY 868 -----GGCTCCGAGCCAGTCACACTGTGGAGGCGTCTCCAGTAAGCAGAGGCC 921
Db 1051 ArgLysSerGluGluLysThrAspThrThrValLysThrMetGluGluTyr 1070
QY 922 GAGAACGAT-----CAGGGGGTGTG-----TCC 945
Db 1071 AsnAsnAspThrAlaProAlaGluAspValIleMetIleHisValProGlnSer 1090
QY 946 CAGGCTGGGAAGATGACAGAGAGTAAGTTGGAGGGAAGCGGCTCAGCACATCAA 1005
Db 1091 LysTrpAspLysAspAspPheGluSerGluGluGlu-----AspValLys 1105
QY 1006 GATGAGGAGCT---GGAGACTTGGCGCCGACCG----- 1035
Db 1106 ThrThrGlnProIleGlnSerValGlyLysProSerIleLysAsnValThrThr 1125
QY 1036 AGCCT-----GAATGTGAGGTAGACCCCAAGCGCTGTAT 1074
Db 1126 LysProSerAlaThrAlaLysTyrThrGluLysGluSerGluGlnProGluLysLeuGln 1145
QY 1075 TGCAATTGCGCCAGCTCACAAACAGGTTTATGATTGCTGTGACCGCTGTGAAGAA 1134
Db 1146 LysLeuProLysGluAlaSerHis-----GluLeuMetGlnHis 1158
QY 1135 TGCTTTTCATGCGGATTTGTGGGCATTTCTGAGCTCGAGGAGGCTTTTGAAGG--- 1191
Db 1159 GluLeuArgSerSerLysGlySerAlaSerSerGluLysGlyArgAlaLysAspArgGlu 1178
QY 1192 ---AATGGGGAAGACTATATCTGCCCAAC----- 1218
Db 1179 HisSerGlySerGluLysAspAsnProAspLysArgLysSerGlyAlaGlnProAspLys 1198
QY 1219 ---TGACCACTTCGACGAGGATGAGATCTATTCAGAACGCGCATCAGCAGGAA 1275
Db 1199 GluSerThrValAspArgLeuSerGluGlnGlyHisPheLysThrLeuSerGlnSerSer 1218
QY 1276 GCTAATGGAGACTGGAGATGCTGTGACCGATGTACAAGTATAGAACAAATAGAG 1335
Db 1219 LysGluThrArg----- 1222
QY 1336 CAGAACTCTAGCGAAGCAAGGATAAGGGGTAGAATTGAGAAAGCTGCAAAATCCAGT 1395
Db 1223 ---ThrSerGluLysHisGluSerValArgGlySerSerAsnLysAspPheThrProGly 1241
QY 1396 GGCAGAGAACTCAAGATCTCCAGCCTGTGATAGAGCGGCTGCTGCCTCAAAATGT 1455
Db 1242 ArgAspLysLysValAsp----- 1247
```

```
QY 1456 ATTGGCCCCGGGTGCTGTCTACGTGGCGCAGCCGACTCGGTGACTGCAATGACTGT 1515
Db 1248 -----TyrAspSerArgAspTyr 1253
QY 1516 ATCTCTAAACACGCGCGCAGCACAATGAGTTTCTAAGCTCAGGTAAGAACAGAGCCA 1575
Db 1254 SerSerSerLysArgArgAspGluArgGlyGluLeuAlaArgArgLysAspSerProPro 1273
QY 1576 AAGCTTAAGAA-----AAGATGAAGATGAAGCCAGAGAGAGCCAGTCTT 1620
Db 1274 ArgGlyLysGluSerLeuSerGlyGlnLysSerLysLeuArgGluGluArg---AspLeu 1292
QY 1621 CCGAAATGCGGTGCTCAGCGCAGGTATTAATAATCTCTCTGTG----- 1662
Db 1293 ProLysLysGlyAlaGluSerLysLysSerAsnSerSerProProArgAspLysLysPro 1312
QY 1663 -----CACAGAGACAGCTCCAGAAAAAAGAGACACACAGTGAAGAGCGGCTG--- 1713
Db 1313 HisAspHisLysAlaProTyrGluThrLysArgProCysGluGluThrLysProValAsp 1332
QY 1714 -----GTGTCCTCTGCGCGGAGT-----GAAGCA 1737
Db 1333 LysIleSerGlyLysGluArgGluLysHisAlaAlaGluAlaArgAsnGlyLysAspSer 1352
QY 1738 CTCGGGAAGGAGCAGCTTGTGAGAGCAGCAGCGCTGCTGGCGGAGGATCACAATTAC 1797
Db 1353 SerGlyGlyLysLeuProCysIleLeuAsnProProAspLeuProMetGluLysGluLeu 1372
QY 1798 AATCAGTAAAGCCAGAGAAAGACACTGCTGCT---CCCTCGCCCTCACTG 1842
Db 1373 AlaValGlyGlnValGluLysSerAlaValLysProLysProGlnLeu 1388
```

Search completed: April 28, 2004, 11:01:34  
Job time : 86 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 28, 2004, 11:00:28 ; Search time 115.5 Seconds  
(without alignments)  
12527.529 Million cell updates/sec

Title: US-09-787-016A-1

Perfect score: 4689  
Sequence: 1 ctcggtgcgcgcgcac.....actcttaagatcatatctcg 2610

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 1138120 seqs, 277189581 residues

Total number of hits satisfying chosen parameters: 2276240

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp  
-DB=/cgn2\_1/USPTO\_spool\_p/US09787016/runat\_28042004\_094331\_27112/app\_query.fasta\_1.2759  
-DB=Published Applications AA -OFMT=fastan -SUFFIX=rapb -MINMATCH=0.1  
-LOOPEL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsu62  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-NCPUL=2000000000 -USER=US09787016 @CGN 1.1.17 @runat\_28042004\_094331\_27112  
-NCPUL=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCKS=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA.\*

1: /cgn2\_6/ptodata/1/pubaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubaa/US10D\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	958	20.4	181	9	US-09-925-301-1169	Sequence 1169, Ap
2	181.5	3.9	268	9	US-09-764-864-1258	Sequence 1258, Ap
3	180.5	3.8	669	10	US-09-823-187-88	Sequence 88, Appl
4	180.5	3.8	669	10	US-09-863-776-53	Sequence 53, Appl
5	180	3.8	19695	15	US-10-084-846A-3	Sequence 3, Appli
6	177.5	3.8	19695	15	US-10-084-846A-7	Sequence 7, Appli
7	176	3.8	3551	12	US-10-263-929-144	Sequence 144, App
8	176	3.7	19725	15	US-10-084-846A-4	Sequence 4, Appli
9	172	3.7	676	10	US-09-823-187-24	Sequence 24, Appl
10	172	3.7	676	10	US-09-863-776-14	Sequence 14, Appl
11	171.5	3.7	555	14	US-10-106-698-5847	Sequence 5847, Ap
12	171	3.6	678	14	US-10-203-708-28	Sequence 28, Appl
13	171	3.6	678	10	US-09-823-187-87	Sequence 87, Appl
14	171	3.6	678	10	US-09-863-776-52	Sequence 52, Appl
15	171	3.6	717	14	US-10-203-708-29	Sequence 29, Appl
16	171	3.6	3262	15	US-10-379-381-4	Sequence 4, Appli
17	170.5	3.6	19695	15	US-10-084-846A-5	Sequence 5, Appli
18	170	3.6	667	10	US-09-863-776-16	Sequence 16, Appl
19	170	3.6	667	10	US-09-863-776-18	Sequence 18, Appl
20	169.5	3.6	19725	15	US-10-084-846A-4	Sequence 4, Appli
21	169	3.6	1404	9	US-09-811-045A-1	Sequence 1, Appli
22	169	3.6	2781	12	US-10-263-929-122	Sequence 122, App
23	169	3.6	19723	15	US-10-084-846A-5	Sequence 5, Appli
24	168.5	3.6	518	9	US-09-925-299-862	Sequence 862, App
25	168.5	3.6	518	9	US-09-925-299-862	Sequence 862, App
26	168	3.6	670	10	US-09-823-187-86	Sequence 86, Appl
27	168	3.6	670	10	US-09-863-776-51	Sequence 51, Appl
28	168	3.6	671	10	US-09-946-374-308	Sequence 308, App
29	168	3.6	671	10	US-09-823-187-85	Sequence 85, Appl
30	168	3.6	671	10	US-09-863-776-20	Sequence 20, Appl
31	168	3.6	671	12	US-10-206-915-346	Sequence 346, App
32	168	3.6	671	12	US-10-199-670-346	Sequence 346, App
33	168	3.6	671	12	US-10-201-858-346	Sequence 346, App
34	168	3.6	671	12	US-10-205-890-346	Sequence 346, App
35	168	3.6	671	12	US-10-208-024-346	Sequence 346, App
36	168	3.6	671	12	US-10-201-853-346	Sequence 346, App
37	168	3.6	671	12	US-10-174-581-346	Sequence 346, App
38	168	3.6	671	12	US-10-176-483-346	Sequence 346, App
39	168	3.6	671	12	US-10-176-749-346	Sequence 346, App
40	168	3.6	671	12	US-10-176-914-346	Sequence 346, App
41	168	3.6	671	12	US-10-176-915-346	Sequence 346, App
42	168	3.6	671	12	US-10-006-485A-308	Sequence 308, App
43	168	3.6	671	12	US-10-013-907A-308	Sequence 308, App
44	168	3.6	671	12	US-10-015-499A-308	Sequence 308, App
45	168	3.6	671	12	US-10-176-484-346	Sequence 346, App

#### ALIGNMENTS

##### RESULT 1

US-09-925-301-1169  
; Sequence 1169, Application US/09925301  
; Patent No US20020052308A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: P106  
; CURRENT APPLICATION NUMBER: US/09925301  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05982  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1694  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1169  
; LENGTH: 181  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-925-301-1169

Alignment Scores: 5.29e-58 Length: 181  
Pred. No.: 181

Score: 958.00 Matches: 179  
 Percent Similarity: 98.35% Conservative: 0  
 Best Local Similarity: 98.35% Mismatches: 2  
 Query Match: 20.43% Indels: 0  
 DB: 9 Gaps: 0

US-09-787-016A-1 (1-2610) x US-09-925-301-1169 (1-181)

QY 998 ACATCAAGATGAGGAGCTGAGACTTG-GCCGACCGAAGCTGAATGTGAGGGTTAC 1056  
 DB 1 ThrSerLysMetArgSerLeuGluThrLeuGlyArgProLysProGluCysGluGlyTyr 20  
 QY 1057 GACCCCAAGCCCTGATGCTATGCTATGCTGCGCAGCTCACAACAACAGCTTATGATTC 1116  
 DB 21 AspProAsnAlaLeuTy-CysileCysArgInProHisAsnAsnArgPheMetileCys 40  
 QY 1117 TGTGACCGCTGTGAAGATGTTTCATGCGCAATGTGTGGCATTCTTCGAGGCTCGAGGG 1176  
 DB 41 CysAspA-gCysGluGluTrpPheHisGlyAspCysValGlyIleSerGluAlaArgGly 60  
 QY 1177 AGCTTTTGGAAAGAAATGGGAAGACTATATCTGCCCAAACTCCACCATCTTGCAGATG 1236  
 DB 61 ArgLeuLeuGluArgAsnGlyGluAspTyrileCysProAsnCysThrIleLeuGlnVal 80  
 QY 1237 CAGGATGAGACTCATTACAAAAGCCAGATCAGCAGGAAGCTAAATGAGACCTGGAGAT 1296  
 DB 81 GlnAspGluThrHisSerGluThrAlaAspGlnGlnGlnAlaLysTTPA-gProGlyAsp 100  
 QY 1297 GCTGATGCCACGATGTACAGTATAGGAACATATAGGACAGCAAGTCTACGGAAGACAA 1356  
 DB 101 AlaAspGlyThrAspCysThrSerileGlyThrileGluGlnLysSerSerGluAspGln 120  
 QY 1357 GCGATAAAGGGTAGAATTGAGAAAGCTCAAAATGCAAGTGGCAAGAGAAACTCAAGATC 1416  
 DB 121 GlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLysLysLysLeuLysIle 140  
 QY 1417 TTCAGCTGTGATAGAGCGCTGGTCCCTCAAAATGATGCTGCGCGGTGCTGCAC 1476  
 DB 141 PheGlnProValileGluAlaProGlyAlaSerLysCysileGlyProGlyCysCysHis 160  
 QY 1477 GTGGCGCAGCCGACTCGCTGCTACTGCTAGTATGCTATCTCAACACGCGCGAGCG 1536  
 DB 161 ValAlaHis-ProThrA-gCysThrAlaValMetThrValSerSerAsnThrProGlnAr 180  
 QY 1537 ACA 1540  
 DB 180 gGln 181

## RESULT 2

US-09-764-864-1258  
 ; Sequence 1258, Application US/09764864  
 ; Patent No. US20020132753A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 ; FILE REFERENCE: PT223  
 ; CURRENT APPLICATION NUMBER: US/09/764,864  
 ; PRIORITY FILING DATE: 2001-01-17  
 ; Prior application data removed - consult PALM or file wrapper  
 ; NUMBER OF SEQ ID NOS: 1792  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 1258  
 ; LENGTH: 268  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: SITE  
 ; LOCATION: (152)  
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 ; NAME/KEY: SITE  
 ; LOCATION: (209)  
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 ; NAME/KEY: SITE

; LOCATION: (264)  
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 ; NAME/KEY: SITE  
 ; LOCATION: (265)  
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 ; NAME/KEY: SITE  
 ; LOCATION: (268)  
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 ; US-09-764-864-1258

## Alignment Scores:

Pred. No.: 0.000459 Length: 268  
 Score: 181.50 Matches: 52  
 Percent Similarity: 43.09% Conservative: 29  
 Best Local Similarity: 27.66% Mismatches: 62  
 Query Match: 3.87% Indels: 45  
 DB: 9 Gaps: 10

US-09-787-016A-1 (1-2610) x US-09-764-864-1258 (1-268)

QY 934 GGGTTGTGTCCTCCAG---GCTGGGAAAGATCAGAGAGAGTAAGTTGGAGGGAAGCGG 990  
 DB 29 GlyValValArgGluCysAlaGlySerGlyAspMetGluGly----- 42  
 QY 991 GCTCAGGACATCAAAAGATGAGCAGCTGGAGACTTGGCCGACCCAAAGCCTGAATGTGAG 1050  
 DB 43 -----AspGlySerAspProGluProAspAlaGlyGluAspSer---LysSerGlu 59  
 QY 1051 GGTACAGCCCAAGCCCTGTATTGCTTTGCTGCGCAGCTCACAACAACAGCTTTATG 1110  
 DB 60 AsnGlyGluAsnAlaProIleTyrCysileCysArgLysProAspIleAsnCysPheMet 79  
 QY 1111 ATTTCCTGTGACCGCTGTGAAGAAATGTTTCATGCGCATTTGTGGCGCATTTCTGAGGCT 1170  
 DB 80 IledGlyCysAspAsnCysAsnGluTrpPheHisGlyAspCysileArgIleThrGluLys 99  
 QY 1171 CAGGAGCGCTTTGGAAAGAAAGGAGAGACTATATCTGCCCAAACTGC----- 1221  
 DB 100 MetAlaLysAlaIle-----ArgGluTrpTyrCysArgGluCysArgGluLys 115  
 QY 1222 ---ACCATTTCTGCAAGTGCAG----- 1242  
 DB 116 AspProLysLeuGluIleArgTyrArgHisLysLysSerArgGluArgAspGlyAsnGlu 135  
 QY 1243 GAGACTCATTCAGAAACCGCAGATCAGCAGGAGACTAAATGGAGACCT---GGAGATGCT 1299  
 DB 136 ArgAspSerSerGluProArgAspGluGlyGlyArgLysArgProValProAspPro 155  
 QY 1300 GATGGCACCGATTGTCAAGTATAGGAACAATAGCAGCAAGTCTAGCGAAGACCAAGGG 1359  
 DB 156 Asp-----LeuGlnArgArgAla\*\*\*SerGlyThrGly 166  
 QY 1360 ATAAAGGCTAGATTTAGAAA---CTGCAATCCAAAGTGGCAAGAGAAACTCAAGATC 1416  
 DB 167 ValGlyAlaMetLeuAlaArgGlySerAlaSerProHisLysSerSerPro----- 183  
 QY 1417 TTCAGCGCTGTGATGATGAGGCGCT 1440  
 DB 184 ---GlnProLeuValAlaThrPro 190

## RESULT 3

US-09-823-187-86  
 ; Sequence 88, Application US/09823187  
 ; Publication No. US20030096952A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Burgess, Catherine  
 ; APPLICANT: Gusev, Vladimir Y  
 ; APPLICANT: Liu, Xiaohong  
 ; APPLICANT: Majumder, Kumud  
 ; APPLICANT: Padigaru, Muralidhar  
 ; APPLICANT: Patturajan, Meera  
 ; APPLICANT: Shinkets, Richard A  
 ; APPLICANT: Spaderma, Steven K

220	AGGATTTCAGGGAAAAGTGTCTCAAGCTTTTCAGTGTGTGGACGAGGTATGCAACGCAAGGC	276
	:::: :::: :::: :::: :::: ::::	
171	ArgArgAlaSerSerAspLeuAspGlnAlaSerValSerProSerGluGluAspSerGlu	190
	:::: :::: :::: :::: :::: ::::	
280	GACCGCGCAATGAGGAGGCACCTTAAGCCCATCAACCCACACGACAAAGAGTTTCAGGAAA	339
	:::: :::: :::: :::: :::: ::::	
191	SerProSerGluSerGlu-----LysThrSerAspGlnAspPhe-----	203
	:::: :::: :::: :::: :::: ::::	
340	ACATGGGGTTTTCGAAGAGCACCATATCGCAAGCGAGAGGGCGGACGAGGACCGGAGGCT	399
	:::: :::: :::: :::: :::: ::::	
204	-----ThrProGluLysLysThrAlaAla-----	211
	:::: :::: :::: :::: :::: ::::	
400	GACCCACTGAGGCGGCCACCCCCACACGACGACGTGGCGCTGTCCCTCGCGCGCCAGTGGG	459
	:::: :::: :::: :::: :::: ::::	
212	-----ArgProProArgGlyProLeuGly-----Gly	221
	:::: :::: :::: :::: :::: ::::	
460	AGGCGAGCCCAAGCGCACTGACGCGGTGGAGCAGTTCTCTGACCACTTCGGCGGCGCGCGGC	519
	:::: :::: :::: :::: :::: ::::	
222	ArgLysLysLysLysValProSerAlaSerAspSerAspSerLysAlaAspSerAspGly	241
	:::: :::: :::: :::: :::: ::::	
520	AGGAGGACATGCTCTGTCTCCTCGAGGAAATCTTGGTGAGCCCACTGCTCTGCCCCGCCACA	579
	:::: :::: :::: :::: :::: ::::	
242	AlaLysGluGluProValVal-----ThrAlaGlnProSer-----ProSerSer	256
	:::: :::: :::: :::: :::: ::::	
580	GACGCGGAGACAGCTCCGAGGGCAGCGTGGAAAGCGCTTCTGACACACGAGAAGCGGCCCC	639
	:::: :::: :::: :::: :::: ::::	
257	SerSerSerSerSerSerSerSerSerSerAspValSerValLysLysProPro	276
	:::: :::: :::: :::: :::: ::::	
640	CAGTCTGCTTCACAGCTGTGAAG-----	663



Db 365 GlySer-----SerGlyGluGluLeuGluAsp 373  
QY 928 GATCAGGGGTTGTCTCCAGGCTGGGAAGATGACAGAGAGTAAGTTGGAGGAAAG 987  
Db 374 GluGluProVal-----LysLysArgSerArgLysAlaArgLysArg 387  
QY 988 GCGGCTCAGACATCAAGATCAGAGCCT---GGAGACTTGGGCGCA-----1032  
Db 388 GlyThrProSerSerSerAspSerGluProGluGlyGluLeuGlyLysLysLys 407  
QY 1033 -----CCGAAGCTGAATGTGAGGGTTACAGCCCAAGCCCTGTATTGCAATTTGCCGC 1086  
Db 408 LeuAlaLysLysSerGlnLeuProGlySerGluSer-----AlaArg 421  
QY 1087 CAGCCTCACAACAACAGGTTTATGATTCTGTGACCGCTGTGAAGAATGGTTTCATGCC 1146  
Db 422 LysPro-----423  
QY 1147 GATTGTGTGGCATTTCTCAGGCTCGAGGAGGCTTTTGGAAAGGAATGGGAAGACTAT 1206  
Db 424 -----GlyGlnLysGluLysArgGlyArg-----431  
QY 1207 ATCTGCCAAACTGCACCATCTGCAAGTGCAGATGAGACTCATTCAAGAAAGCGGAGAT 1266  
Db 432 -----Pro 432  
QY 1267 CAGCAGGAAGCTTAATGGAGACCTGGAGATGCTGTGCAACCGAATGTACAAGTATAGGA 1326  
Db 433 AspGluLysProArgAlaArgProValLysValGluArgThr-----446  
QY 1327 ACAATGAGCAGAGTCTACGAGAGCAAGGATTAAGGTAAGTATGAGAAGCTGCA 1386  
Db 447 -----ArgLysArgSerGlyLysLeuSerLeuGluArgLysGlyLysLysLys 463  
QY 1387 AATCCAAGTGGCAAGAACTCAAG-----ATCTTCCAGCCTGTG 1428  
Db 464 GluProSerValGluGluArgLeuGlnLysLeuHisSerGluLysPheAlaLeuLys 483  
QY 1429 ATAGAGCGCTGTGTCTCAAAATGATTATGGCCCC-----GGTCTGTCTAC 1476  
Db 484 ValAspAsnProAspValArgLysCysLeuSerAlaLeuGluGluLeuGln 503  
QY 1477 GTGGCG-----CAGCCGACTCGGTG-----1497  
Db 504 ValThrSerGlnLeuGlnLysAsnThrAspValValAlaThrLeuLysLysLysLys 523  
QY 1498 ---TACTGTCAGTAACTGCTATCTCTCAAAACAGCCGCA-----1533  
Db 524 ArgTyrLysAlaAsnLysAspValMetAlaLysAlaAlaGluValTyrThrArgLeuLys 543  
QY 1534 -----CCGACATGAAGTTTCTAAGCTCAGCTAAAGAA 1566  
Db 544 SerArgValLeuGlyProLysValGluAlaLeuGlnLysValAsnLysAlaGlyAlaGlu 563  
QY 1567 CAGAAGCAAAAGCCTAAAGAAAGATGAAGTGAAGCCAGAGAGCCAGTCTTCGAAA 1626  
Db 564 LysGluArgAlaAspAsnGluLysLeuGluGluGlnPro-----576  
QY 1627 TGGGCTCAGGAGTATTAAATCTTCTGTGCAC-----AAGAGACAGCT 1677  
Db 577 ---GlyGluGlnAlaProArgGluLeuAlaGluAspGluProSerThrAspArgSerAla 595  
QY 1678 CCAGAAAAAAGAGACACACATGACAGAGCAGTGTGTGCTGCTGCGGAGTGAAGCA 1737  
Db 596 ProValAsnGlyGluAlaThrSerGlnLysGlyGluAsnMetGluAspArgAlaGlnGlu 615  
QY 1738 CTCGGGAAGGAGCAGTGTGTGAGAGCAGCAGCCGCTGCTGGGCGGAGCGATCAC-----1791  
Db 616 AspGlyGlnAspSer-----GluAspGlyProArgGlyGlySerSerGluGluLeuHis 633  
QY 1792 -----AATTACAATGCAGTAAAGCA-----GAA 1815  
Db 634 AspSerProArgAspAsnSerAspProAlaLysProGlyAsnGluArgGlnAspHisGlu 653

QY 1816 AAGACTGCTGCTCCTCCGCGCTCA 1839  
Db 654 ArgThrArgLeuAlaSerGluSer 661  
RESULT 5  
US-10-084-846A-3  
; Sequence 3, Application US/10084846A  
; Publication No. US20040006026A1  
; GENERAL INFORMATION:  
; APPLICANT: WEITNAUER, GABRIELE  
; APPLICANT: MUELENWEG, AGNES  
; APPLICANT: TREFFZER, AXEL  
; APPLICANT: BECHTOLD, ANDREAS  
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES  
; FILE REFERENCE: 1974-005  
; CURRENT APPLICATION NUMBER: US/10/084,846A  
; PCT FILING DATE: 2003-02-25  
; PRIOR APPLICATION NUMBER: PCT/EP01/09815  
; PRIOR FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: DE 101 09 166.4  
; PRIOR FILING DATE: 2001-02-25  
; NUMBER OF SEQ ID NOS: 120  
; SOFTWARE: PatentIn Ver. 3.2  
; SEQ ID NO 3  
; LENGTH: 19695  
; TYPE: PRT  
; ORGANISM: Streptomyces viridochromogenes  
; FEATURE:  
; OTHER INFORMATION: Protein 1: amino acid sequence encoded by coding strand 1.  
; OTHER INFORMATION: Start codon: gga, Start position: nucleotide 1.  
US-10-084-846A-3  
Alignment Scores:  
Pred. No.: 0.00178 Length: 19695  
Score: 180.00 Matches: 158  
Percent Similarity: 30.64% Conservative: 77  
Best Local Similarity: 20.60% Mismatches: 259  
Query Match: 3.84% Indels: 273  
DB: 15 Gaps: 33  
US-09-787-016A-1 (1-2610) x US-10-084-846A-3 (1-19695)  
QY 209 TGGTTGTATTATTAGGATTTTCAGGAAAGTGTCCAAAGCTTTCAGTGTGTGAGCAGGTATGG 268  
Db 10248 TTPTyTPheValAlaAspArgGlyProLeuSerAlaArgSerProArgGlyGluValHis 10267  
QY 269 AGCAAAAGCGACCGACCAATGAGGAGGACCTAAGCCCATCAACCCACCCAGCAAG 328  
Db 10268 SerThrProIle-----GlnLeuProProProSerProSerArg 10281  
QY 329 AGT-----TCAGAAAACATGGGGTTTTCGAGAGGACCACTATCGCCAAAGCAG 376  
Db 10282 SerProAsnAlaIleSerGlyArgProAlaCysSerArgSerThrSerProSer--- 10300  
QY 377 AGGGCGCAGGGGACGCGG-----AGGCTGACCCACTCGGAC 412  
Db 10301 -----GlyGlyArgValCysTPProCysAlaAlaCysAlaProAsnTrpSer 10317  
QY 413 CGCCACCCCAACAGCAGCAGCTGGGCTGTCTCCCTGGGCGCAGTGGGAGGAGCCCAAGC 472  
Db 10318 ArgCysArgArgArgProAlaSerThrThrProGlyAlaProGlyProCysThrThr 10337  
QY 473 GCATGAGCGGTGGAGCAGTCTCTGACCATGCGCGCGCGCGCGGAGGAGGA-----526  
Db 10338 ProSerGlyArgTrpSerProThrProThrArgProArgSerAlaSerSerThrSer 10357  
QY 527 -----GCATGCTGTCTCTCTCTGAGG 547  
Db 10358 TTPThrSerProArgTrpProGlyTrpArgAlaIleArgCysCysAlaProTrpArg 10377  
QY 548 ATTCTGTGAGCCCAAGTCTCTGCGCCGACACAGCCCG-----AGACAGCTTCGAGG 601  
:::|||||

Db 10378 ProTpsSerGlyArgThrSerSerProProGlyThrAlaTrpCysSerArgProProArg 10397  
QY 602 GCA-----GCGTGAAGAGCGCTCTGAGACAGAGCGCCCGCCAGTCTGCTCCACAG 655  
Db 10398 AlaArgProGlyTrpProGlyThrValThrAlaArgCysThrProMetProSerPro-Se 10417  
QY 656 CTGTGAAGAAACGACCGCTCTTCTGAAGAGTGAAGAGGGGATGACCCAGCATGACA 715  
Db 10417 rProAlaAlaValAlaSerSerThrTrpAlaSerThrSerThrThrArgProArgThrTh 10437  
QY 716 CTCCCGATGATGACAGCGCTGCTGACCTTGAAGAGCTTGAAGATCCCTTCGCGAGCA 775  
Db 10437 rValSerGlyProSerArgSerProThrThrThrGlyArgThrSerArgSerProArgProPr 10457  
QY 776 AGCGGGA----- 782  
Db 10457 oThrGlySerThrProProSerGlyThrArgProGlyProSerProProCysGlyArgAl 10477  
QY 783 -----ACAGGCCCACTGACAGGCCCTCGAA----- 809  
Db 10477 aThrProCysCysThrThrSerHisCysThrAlaProArgArgTrpTrpAlaSerSerAl 10497  
QY 809 ----- 809  
Db 10497 aGlySerSerThrThrSerThrGlyProAlaArgTrpSerValSerSerAlaArgThrCy 10517  
QY 809 ----- 809  
Db 10517 sProSerThrTrpGlySerSerSerArgCysAlaArgAlaTrpSerSerGlyProAr 10537  
QY 810 -----AGGATCCAGAGTGCCTCGCGAA-----GAGCGCGCGGAGG 847  
Db 10537 gProGlySerThrGlyThrArgSerProSerSerThrAlaArgTrpSerSerThrGlySe 10557  
QY 848 AGGTCCCGCGAGAC-----TGCGGCTCCGAGGCGCAGTGCACACTGTGAGG 895  
Db 10557 rGlyThrSerArgProArgSerProAlaCysGlySerArgThrArgSerThrGlyGlyGl 10577  
QY 896 GCGTCTCGCCAG-----TAAGCAGGAGCCGAGACAGTACAGGGGTTGTGCCAGG 949  
Db 10577 yGluProAlaGluProProAlaAlaAlaGlyArgAlaArgProGlyAlaGlyAlaProGl 10597  
QY 950 CTGGAAAGATGACAGAGAGTAAGTTGGAGGGAAGCGGCTCAGCAGATCAAGA-- 1007  
Db 10597 y--ArgAlaHisArgPro-----GlyGlyHisArgArgAl 10609  
QY 1008 -----TGAGGAGCTGGAGACTTGGCGAGCGAGCTGAATGTGAGGTTAGACCCCA 1063  
Db 10609 aLeuProGlyAlaAlaArgSerAlaPro--ArgArgLeuGlyLeuArgProSerAlaArg 10628  
QY 1064 ACGCCCTGTATTGCTTGGCCGAGCTCACAACACAGTTTATGATTTGCTGTGACC 1123  
Db 10629 ProProGlnArgAlaGlyArgAlaAlaArgGlyValGly-----AlaLeuAla 10645  
QY 1124 GCTGTGAAGATGCTTTCATGCGGNTTGTGGGCTCTTCTGAGGCTC----- 1171  
Db 10646 ArgLeuArgSerGlyHisLeuAlaValProArgProAlaProArgA-gProAlaGlyAsp 10665  
QY 1172 -----GAGGAGGCTTTTGGAAAGGAATGGGAAGACT 1204  
Db 10666 GlyArgAspAlaArgArgArgGlyProAspGlyGlySerArgAlaGlyArgGlyArgAla 10685  
QY 1205 ATA----- 1207  
Db 10686 ValArgProAlaGlyArgLeuValArgArgCysAlaLeuHisProGlyProGlyThrArg 10705  
QY 1208 -----TCTGCCCAACTCCACATCTCTCAAGATGCGAGGATGAGACTC 1249  
Db 10706 AspGlyProGlyProAlaSerAlaArgGlyAlaPro-AlaAlaAlaAla----- 10721  
QY 1250 ATTCAAGAACCGCATCAGCAGAGCTAATGAGACCTGGAGATGCTGATGCGACCG 1309  
Db 10722 -----GlyArgGlnAlaGlyAla--ValHisAlaTrpArg-LeuGlyGlyAlaA 10737

QY 1310 ATTGTACAGTATAGGACAAATAGACGAGAGTCTAGCGAGAGCAACCAAGGGATAAAGGCTA 1369  
Db 10737 rGArgAlaGlyProValAlaLeuAspArgA-gArgThrArgGlyProGlyArgAlaGlyA 10757  
QY 1370 GAATTGACAAAGCTGCAAAATCCAAAGTGGCAAGAAACAACTCAAGATCTTCCAGCCTGTGA 1429  
Db 10757 rGValHisArg-----GlyProAsnGlyGlnAlaHisLeuArgThrAlaArgSerAlaA 10775  
QY 1430 TAGAGCG-----CCTGTGCTCAAAATGATTGGC--CCCGGCTGCTGTC 1474  
Db 10775 laProAlaAlaLeuProArgProGlyAlaAlaGlyGlyArgSerGlyArgAlaA 10795  
QY 1475 ACGTGGCGCAGCCGACTCGGTGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1534  
Db 10795 laAlaAlaAspPro----- 10799  
QY 1535 CGACAATGAAGTTTCTAGCTCAGTAAAGACACAGACCCAAAGCCTAAAGAAAGATGA 1594  
Db 10800 -----GlyProGlyArgProAlaAlaProAlaProArgArg-- 10812  
QY 1595 AGATGAAGCCAGAGAGCCAGTCTTCGAAATGCGGTCTCAGGCAGGTATTAATAATCT 1654  
Db 10813 -----LeuProArgProGlyLeu----- 10818  
QY 1655 CTTCTGTGCACAGAGACAGCTCCAGAAAAGAGACACACAGTGAAGAGGCGAGTGG 1714  
Db 10819 -----ArgArgProArgProArgArgGly----- 10826  
QY 1715 TGGTCCCTCGCGAGTGAAGCACTCGGAGAGAGCAGCTTGTGAGACAGCAGCCCT 1774  
Db 10827 -----ProGlyArgAlaValLeuProGlyHisProAlaAspArgGlnGlnAlaAlaArgA 10845  
QY 1775 CTTGGCGCAGCGATCACAATTAACATGCAAGTAAAGCCAGAAAGACTGCTGCTCCCTCC 1834  
Db 10845 laGlyGluAlaGlyGlyArgGlyArgMetThrGlnPro-----ProProSerP 10861  
QY 1835 CTTCACTGTTGTAATAAGTATGATCATCTAGGCGGTGGCTCTGAGCCCTCCCGTT 1894  
Db 10861 roThrMet-----CysVal----- 10865  
QY 1895 CTTCTGATGATGATCCCTCCCTG-----GCCTGTCCAGACTG--GAGTTGAGCTT 1945  
Db 10866 -----LeuProTrpIleHisLeuCysAlaSerIleAspGlyValTyrGlyA 10881  
QY 1946 TGTGTTAAGCTGATCAGCA-----CACCGCTGCACCATCAGCGGAGGAGCAGCCCA 1999  
Db 10881 rGysCysValAspAspSerMetTyrHisAsnGluLeuTyrGluSerGluAlaGluProV 10901  
QY 2000 TGTCCAGGATG 2010  
Db 10901 alPheLysLeu 10904

RESULT 6

US-10-084-846A-7  
; Sequence 7, Application US/10084846A  
; Publication No. US20040006026A1  
; GENERAL INFORMATION:  
; APPLICANT: WEITNAUER, GABRIELE  
; APPLICANT: MUHLERWEG, AGNES  
; APPLICANT: TREFFZER, AXEL  
; APPLICANT: BECHTHOLD, ANDREAS  
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES  
; FILE REFERENCE: 1974-005  
; CURRENT APPLICATION NUMBER: US/10/084,846A  
; CURRENT FILING DATE: 2003-02-25  
; PRIOR APPLICATION NUMBER: PCT/EP01/09815  
; PRIOR FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: DE 101 09 166.4  
; PRIOR FILING DATE: 2001-02-25  
; NUMBER OF SEQ ID NOS: 120  
; SOFTWARE: PatentIn Ver. 3.2  
; SEQ ID NO 7





QY 1659 TGTGCAAGAGAC-----CAGCTCCAGAAAAAAGAGCAGCAGTGAAGAA 1706  
 Db 3912 laCyethrArgGlnCysGlyThrProSerThrSerLeuThrSerMetProArgWetProI 3932  
 QY 1707 GGCAGTGGTGGTCCCTCGCGGAGTGAAGCAGCTCGGGAAGGAAGCAGCTTGTGAGAGCAG 1766  
 Db 3932 leserGlyTrpLeuMetMetGlyTrpLeuHisSerGlyThrGlyProLeuValArgThra 3952  
 QY 1767 CACGC 1771  
 Db 3952 rgarg 3953

RESULT 7  
 US-10-263-929-144  
 ; Sequence 144, Application US/10263929  
 ; Publication No. US20040067535A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kim, Jaeseob  
 ; APPLICANT: Galant, Ron  
 ; TITLE OF INVENTION: Alzheimer's Disease Linked Genes  
 ; FILE REFERENCE: LSD-07417  
 ; CURRENT APPLICATION NUMBER: US/10/263,929  
 ; CURRENT FILING DATE: 2002-10-03  
 ; NUMBER OF SEQ ID NOS: 213  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 144  
 ; LENGTH: 3551  
 ; TYPE: PRT  
 ; ORGANISM: Mus musculus  
 US-10-263-929-144

Alignment Scores:  
 Pred. No.: 0.00216 Length: 3551  
 Score: 176.00 Matches: 187  
 Percent Similarity: 30.63% Conservative: 93  
 Best Local Similarity: 20.46% Mismatches: 304  
 Query Match: 3.75% Indels: 332  
 DB: 12 Gaps: 38

US-09-787-016A-1 (1-2610) x US-10-263-929-144 (1-3551)

QY 248 TCAGTGTGGAGCAGGTATGGACAAAGGCGACCCGACGAATGAGGAGGACCTAAGG 307  
 Db 1761 SerValValGluLysProValThrArgLysSerGluArgLleAspArgGluLysLeuLys 1780  
 QY 308 CCATCAAAACCCACCA----- 322  
 Db 1781 ArgSerSerProArgGlyGluAlaGlnLysLeuLeuLysMetGluAlaGlu 1800  
 QY 323 -----GCAAGACTTCAGG-----AAACATGGGTTTTTCGAA 355  
 Db 1801 LysIleThrArgThrAlaSerLysSerGlyAspThrGluHisProGluProSer 1820  
 QY 356 GGACCACTATCGCCAGCGAGAGGGGCGCAGG-----GAC 390  
 Db 1821 LeuProLeuSerArgSerArgArgAsnValArgSerValTyraThrMetThrAsp 1840  
 QY 391 GCGGAGGCT-----GACCACTGGAGCCGCCA----- 417  
 Db 1841 HisGluSerArgSerProAlaLysGluProValGluGlnProArgValThrArgLysArg 1860  
 QY 418 -----CCCCCAGCAGCAGCAGCTGGGCGCTGTCCTG 447  
 Db 1861 LeuGluArgGluLeuGlnGluAlaValProPro-----ThrThr 1874  
 QY 448 CGCGCAGTGGGAGGAGCCAGCGCAGCTGAGCGGTGGAGCAGTTCCTGACCATTGGG 507  
 Db 1875 ProArgArgGlyArgProProLysThrArgArgAlaGluGlu----- 1889  
 QY 508 CGCGCGCGCGGAGGAGCAGCATGCTCTCCCTGGAGGATCTGCTGAGCCACGCTCC 567  
 Db 1890 AspGlyGluHisGluArgLysGluProAlaGluThrProArgProAlaGluGlyTrpArg 1909

QY 568 TGCCCGCCACAGACGCGCGAGACAGCTCCGAGGCG----- 603  
 Db 1910 SerProArgSerGlnLysSerAlaAlaAlaAlaGlyProGlnGlyLysArgGlyArgAsn 1929  
 QY 604 -----AGCTGGAAACGGCTTCTGAG-----ACGAGAGCGGC--- 636  
 Db 1930 GluGlnLysValGluAlaAlaAlaGluAlaGlyAlaGlnAlaSerThrArgGluGlyAsn 1949  
 QY 637 CCCAGTCTGCTTCCACAGCTGTGAAGGAACGACGACCTCTTCTGAAAAAGTGAAGGA 696  
 Db 1950 ProLysSerArgGly-----GluArgGluAlaAlaSerGluProLysArg--- 1964  
 QY 697 GGGGATGACCAACGATGACACCTCCGAT---AGTCACAGCGATGCGCTGACCTTGAAGAG 753  
 Db 1965 ---AspArgArgAspProSerThrAspLysSerGlyProAspThrPheProValGluVal 1983  
 QY 754 CTTTCAGATCGCTTCCAGGACCGGGAACGAGAGCCCACTGAGAGCGCCCTCGAAAGGG 813  
 Db 1984 Leu-----GluArgLysProGluLysThrTyrlLysSer 1995  
 QY 814 ATCCAGAGTCCCTCGCGAAGAACCGC----- 840  
 Db 1996 LysArgGlyArgAlaArgSerThrArgSerAlaMetAspArgAlaAlaHisGlnArgSer 2015  
 QY 841 -----CGGAGGAGGGTCCCGCCCGAG 861  
 Db 2016 LeuGluMetAlaAlaArgAlaAlaGlyGlnAlaAlaAspLysGluAlaGlyProAlaAla 2035  
 QY 862 ACTGTGGGCTCCGAGGCGCAGTGTGAGGGCGTCTGCGGCGCTCCGCCAGTAAGCAG----- 915  
 Db 2036 AlaSerProGlnGluSerGluSerProGlnLysGlySerGlySerSerProGlnLeuAla 2055  
 QY 916 -----GAGCCCGAGAACGATCAGGGGTTGTCTCCAGGCT 951  
 Db 2056 AsnAsnProAlaAspProAspArgGluAlaGluGluGluSerAlaSerAlaSerThrAla 2075  
 QY 952 GCGAAAGATGAC-----AGCAGCAGTAGTTGAGGGGAAAGCGCGCTCAGAC 999  
 Db 2076 ProProGluGlyThrGlnLeuAlaArgGlnIleGluLeuGlu---GlnAlaValGlnAsn 2094  
 QY 1000 ATC---AAAGATGAGGACCTGGAGACTTGGGC----- 1029  
 Db 2095 IleAlaLysLeuProGluProSerAlaAlaAlaAlaSerLysGlyThrAlaThrAlaThr 2114  
 QY 1030 -----CGACCGAAGCCTGAATGTGAGGTTACGACCCCGCCAGCCCTGTATTGC 1077  
 Db 2115 AlaAlaSerGluGluProAlaProGlu---HisGlyHisLysPro----- 2128  
 QY 1078 ATTTCCCGCCAGCTCACACACAGGTTTANGATTGCTGTGACCCGCTGTGAAGAATGG 1137  
 Db 2129 -----AlaHisGlnAlaSerGluThr 2135  
 QY 1138 TTTGATGCGGATTGTGTGGC-----ATTTCGAGGCTCGAGGGAGGCTTTTGAAAGG 1191  
 Db 2136 GluLeuAlaAlaAlaIleGlySerIleIleSerAspAlaSerGly----- 2150  
 QY 1192 AATGGGAGAGCTATATCTGCCCA----- 1215  
 Db 2151 GluProGluAsnPheSerAlaProProSerValProProGlySerGlnThrHisProArg 2170  
 QY 1216 -----ACTGCACCATTCGCAAGTCAGGATGAG 1245  
 Db 2171 GluGlyMetGluProGlyLeuHisGluAlaGluSerGlyIleLeuGluThrGlyThrAla 2190  
 QY 1246 ACTCATTTGAGAA-----ACGCCAGATCAG 1269  
 Db 2191 ThrGluSerSerAlaProGlnValSerAlaLeuAspProProGluGlySerAlaAspThr 2210  
 QY 1270 CAGGAAGCTTAATGAGACCTGGAGATCTGATGCCACCGATTGTACAAGTATAGGAACA 1329  
 Db 2211 LysGluThrArgGlyAsnSerGlyAspSer-----Val 2221  
 QY 1330 ATAGAGCAGAGCTCTAGC-----GAAGACCGAGGGATAAAG 1365

```

Db      2222 GlnGluAlaLysGlySerLysValGluValThrProProArgLysAspLysGlyArgGln 2241
QY      1366 GGTAAATGAGAAGCTCAAAATCCAACTGCGCAAGAGAAACTCAAGATCTTC----- 1419
Db      2242 LysThrThrArgArgLysArgAsnAlaAsnLysValValAlaLalleThrGluThr 2261
QY      1420 -----CAGCCTGTGATAGAGGCGCTGGTGGCTCAAAATGATTGGC 1461
Db      2262 ArgAlaSerGluAlaGluGlnThrGlnSerGlnSerProAlaAlaGluAlaThr--- 2280
QY      1462 CCGGGGTGCTGTCACGTGGCGGCGCCGACTCG----- 1494
Db      2281 -----AlaAlaThrProGluAlaProGlnGlnLysGlnSerGluLys 2295
QY      1495 -----GTGTACTGCAGTAATGACTGTATCTCTCAACACGCCCGCAGCGACA 1539
Db      2296 ProProSerProProAlaGluCysThrPheAspProSerLysThrProProAlaGluSer 2315
QY      1540 ATCAAGATTTCTAAGCTCAGGTAAAGAACAGACCAAGCCAAAGCCTAAAGAAAGATGAAGATG 1599
Db      2316 LeuSerGlnGluAsnSerAlaAlaGluLysThrProCysLysAlaProValLeuProAla 2335
QY      1600 AAGCCA-----GAGAAGCCAGTCTTCGGAATGCGGTGCTCAGCAGGTATTAAATC 1653
Db      2336 LeuProProLeuSerGlnProAlaLeuMetAspAspGlyProGlnAlaArgPheLysVal 2355
QY      1654 TCTTCTGTG----- 1662
Db      2356 HisSerIleGluSerAspProValThrProProSerAspSerGlyIleProProPro 2375
QY      1662 ----- 1662
Db      2376 ThrIleProLeuValThrIleAlaLysLeuProProProValIleProGlyValPro 2395
QY      1663 CACAAGAGCCAGCTCCAGAAAAAAGAGACACACAGTGAAGAGCAGTGGTGGTCCCT 1722
Db      2396 HisGlnSerProProProLysValThrGluThrIleThrLys----- 2409
QY      1723 GCGCGGAGTGAACACTCGGAGAGGAGCAGCTGTGTGAGAGCAGCAGCGCGTGGGGG 1782
Db      2410 -----GlnGluLysProArgAlaGlnSerThrProSerProAla 2422
QY      1783 AGCGATCAATTAATGATGATTAAGCCAGAAAGACACTGCTGCTCCCTCGCGCTACATG 1842
Db      2423 LeuProProAsnThrLysAlaSerAspMetAspThrSerSerSerThrLeuArgLysIle 2442
QY      1843 TGTATAAATGATGATGATACCTAGGCGTGGCTCTCGACCCCTCCCGTCTCTTCTGG 1902
Db      2443 LeuMetAspProLysTyrValSerAlaThrGlyValThrSerThrSerValThrSerAla 2462
QY      1903 ATAGCCATCCCTGGGCGCTGTCAGGACTGGGAGTGGCAGCTTGTGTTAAGCTGATCAC 1962
Db      2463 IleAlaGluProValSerAlaProCysLeuGlnGluAlaProAlaProProCysAspPro 2482
QY      1963 AGACACCG-----GCTGCACCATCAGCGGGAAGCAGAGCC 1997
Db      2483 LysHisProProLeuGlnGlyValSerAlaAlaValProAsnAlaAspThrGlnAla 2502
QY      1998 CAT-----GTCCAGGATGCCCTCGCTCGCTGTGTCATCCCTAGT--- 2039
Db      2503 SerGluValProValAlaAlaAspLysGluLysValAlaProValIleAlaProLysIle 2522
QY      2040 -----CTGTACAGACTTCTGTCACTGTT-----TTCCAA 2069
Db      2523 ThrSerValIleSerArgMetProValSerIleAspLeuGluAsnSerGlnLysIleThr 2542
QY      2070 AGCTGTAAACCTCACTGGTGAAGCTTCACTTAATGATTAATCTTTAATCTCTGTTTC 2129
Db      2543 LeuAlaLysProAlaProGlnThrLeuThrGlyLeuValSerAlaLeuThrGlyLeuVal 2562
QY      2130 ACTCTCAGGCTCGGTAAAGTATTGTATTCTCTTCATCCCACTGATTCATAGCCACA 2189

```

```

Db      2563 AsnValSerLeuValProValAsnAlaLeuLysGlyProValLysGlySerValAlaThr 2582
QY      2190 CTGCCCGCGCAGCGCACATCCACCCCTGCTGTCACATGATGTTG 2231
Db      2593 LeuLysGlyValGlyGluHisProTrpTrpAlaArgAspLeu 2596
RESULT 8
US-10-084-846A-4
; Sequence 4, Application US/10084846A
; Publication No. US20040005026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES
; APPLICANT: TREPFER, AXEL
; APPLICANT: BECHTOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084, 846A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCI/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: Patent In Ver. 3.2
; SEQ ID NO 4
; LENGTH: 19725
; TYPE: PRT
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: Protein 2: amino acid sequence encoded by coding strand 1.
; OTHER INFORMATION: Start codon: gat, Start position: nucleotide 2.
US-10-084-846A-4
Alignment Scores:
Pred. No.: 0.00337 Length: 19725
Score: 176.00 Matches: 90
Percent Similarity: 33.24% Conservative: 25
Best Local Similarity: 26.01% Mismatches: 123
Query Match: 3.69% Indels: 108
DB: 15 Gaps: 22
US-09-787-016A-1 (1-2610) x US-10-084-846A-4 (1-19725)
QY      1246 TCTCATCTGCACCTTGACAGATGGTGAGTTGGCAGATATAGTCTTCCCATCTCTTT 1187
Db      10010 SerHisProGly-----HisArgArgProAlaSerArg 10020
QY      1186 CCAAAGCCTCCCTCGACCTCAGAAATGCCACACAAATGCCATGAAACCACTTCTTAC 1127
Db      10021 AlaProGlyGlyGlnArgProGlyArgGlyProArgProArgArgGlyArgAlaArg 10040
QY      1126 AGCGCTCACGCAATATCAAACTGTTGTTGAGGCTGGGGCAATGCAATACAGG 1067
Db      10041 ProGlyArgThr-----AlaGlyArgArgAlaArgAlaGly 10052
QY      1066 CTTTGGGCTGTAAACCTCACATTCAGGCTTGGTGGCCCAAGTCTCCAGGCTCTCAT 1007
Db      10053 Arg---GlnArgGlyProThrAlaArgAlaArgGluValGlyLeuArgTrpProArg 10071
QY      1006 CTTTGAATGCTGAGCGCGCTTCCCTCCAACCTACTCTCTGTCATCTTCCAGCT 947
Db      10072 Arg---CysProSerProSer-----SerThrThr 10080
QY      946 GGGACACAACCCCTGATCTGTTCTCGG-----GCTCTCTCTTACTGG 905
Db      10081 GlyThrThrSerProSerArgSerSerAsnGlySerAlaProProThrProAlaAlaTrp 10100
QY      904 GCAGAGCGCCCTCCACATGTCACGTGCGCTCGAGCCACAGTCTCGGGGGACCTCTCT 845
Db      10101 ThrSerValProAlaAlaAla-----ArgSerSerArgThrCysAlaGlyProPro 10117
QY      844 CCGGGGCTTCTTCGCGAGGAGCTCTGATCCCTTTCCAGGGGCTCTCTCAGTGGGCTCT 785

```







```

Db      3 ArgSerGluPheProGly-----ArgProThrArgProAlaValThrAla 17
Qy      247 TTCAGTGTGGACGAGTATGACGACAAAGCGGACCGAGCAATGAGGCGCACCTAAG 306
Db      18 ThrAlaAlaSerAspArgMetGluSerAspSerAspSerAspSerAspSer 37
Qy      307 GCCATCAAAACCCACGACGAAAGAGTTCAGGAAACATGGGGTTTCGAAAGGACCACTATC 366
Db      38 GlyLeuLys-----ArgLysThrProAlaLeu-----LysMetSerVal 50
Qy      367 GCCAAGCAGAGGGCGCGAGGACCGCGAGGCTGACCCACTCGAGCCGCCACCCACACAG 426
Db      51 SerLysArgAlaArgLysAlaSerSerAspLeuAspGlnAlaSerValSerProSerGlu 70
Qy      427 CAGCAGCTGGGCGCTCTCCCTCGCGCGCAGTGGAGGCGAGCCCAACCGCAGCTGAGCGCGTG 486
Db      71 GluGluAsnSerGluSerSerGluSerGluLysThrSerAspGlnAspPheThrPro 90
Qy      487 GAGCAGTTCCTGACCATTT---CGCGGGCGCCCGCGC-----AGGAGGAGC 528
Db      91 GluLysLysAlaAlaValArgAlaProArgArgGlyProLeuGlyGlyArgLysLysLys 110
Qy      529 ATGCGCTGTCCTCGGAGGATTCTGGTGAGCCCGACGTCCTGCG----- 570
Db      111 LysAlaProSerAlaSerAspSerAspSerLysAlaAspSerAspGlyAlaLysProGlu 130
Qy      571 CCGGCCACAGACCGCGAGACGCTCCGAGGCGGAGCGTGGAAAGCGCTTCT 621
Db      131 ProValAlaMetAlaArgSerAlaSerSerSerSerSerSerSerSerSerSer 150
Qy      622 -----GAGACGAGAGCGCGCCCGAGTCTGCTCCACAGCTGTGAAG----- 663
Db      151 AspValSerValLysLysProProArgGlyArgLysProAlaGluLysProLeuProLys 170
Qy      664 -----GAACGACGAGCTCTTCTGAAAAGGTGAAAGGAGG 699
Db      171 ProArgGlyArgLysProLysProGluArgProProSerSer----- 184
Qy      700 GATGACCAAGTACACCTCCGATAGTGCAGAGTGGCGCTG---ACCTTGAAGAGCTT 756
Db      185 -----SerSerSerAspSerAspSerAspLysGluValAspArgLysSerGluTrp 200
Qy      757 CAGAAATCGC-----CTTCGAGGAGCGCGGAACAGGAG 789
Db      201 LysArgArgAspGluAlaArgArgGluLeuGluAlaArgArgArgGluGlu 220
Qy      790 CCACGTGAGAGGCGCTGAAAGGATCCAGAGTCCGCTCGGAGAGAGCGCGCGAG--- 846
Db      221 GluGluLeuArgArgLeuArgGluGluGluLysGluLysGluArgArgGluArg 240
Qy      847 -----GAGGGTCCCGCGAG-----ACTGTGGGCTCCGAGGCCAGTGCACACTGTGGAG 894
Db      241 AlaAspArgGlyGluAlaGluArgGlySerGlySerGlyAspGluLeuArgGlu 260
Qy      895 GCGCTCTCCCTAGTAAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTCCCGAGGCTGGG 954
Db      261 AspAspGluProValLysLysArgGlyArgLysGlyArgGlyProProSerSer 280
Qy      955 AAGAT---GACAGACAGTAACTAGTTCGGGAAAGCGGCTCAGGACATCAAGATGAG 1011
Db      281 SerAspSerGluProGluAlaGluLeuGluArgGluAlaLysSerAlaLys----- 298
Qy      1012 GAGCCTGGAGACTTGGCGCGACCGACGACCTGAATGTGAGGGTTTACGACCCCAACGCCCTG 1071
Db      299 -----LysProGlnSerSerSerThrGluPro----- 307
Qy      1072 TATTCATTTCGCGCAGCGCTCAACACACAGGTTTATGATTTGCTGTGACCGCTGTGAA 1131
Db      308 -----AlaArgLysProGly----- 312
Qy      1132 GAATGTTTTCATGCGGATTGTGTGGGCAATTTCTCAGGCTCGAGGAGGCTTTTGGAAAGG 1191
Db      312 ----- 312

```

```

Qy      1192 AATGGGAGAGACTATATCTGCCAAACTGCACCAATTCGACAGTGCAGAGTGCAGACTCAT 1251
Db      313 -----GlnLysGluLysArg 317
Qy      1252 TCAGAAAGCGCAGATCAGCAGAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGAT 1311
Db      318 ValArgProGluGluLysGlnAlaLys-----ProValLysValGluArgThr--- 334
Qy      1312 TGTACAATATAGGAACATAGACGACGAGTCTTAGCGAAGCAAGGATTAAGGGTGA 1371
Db      335 -----ArgLysArgSerGluGlyPheSerMetAspArgLys 346
Qy      1372 ATTGAGAAGCTGCAATCCAGTGGCAGAGAACTCAAG----- 1413
Db      347 ValGluLysLysLysGluProSerValGluGlnLysLeuGlnLysLeuHisSerGluLe 366
Qy      1414 ATCTTCCAGCGCTGTAGTAGGCGCCTGGTGCCTCAAAATGATTTGGCCCGGCTGTGT 1473
Db      367 LysPheAlaLeuLysValAspSerProAspValLysArgCysLeuAsnAla----- 383
Qy      1474 CAGTGGCGGACGCGGACTCGTGTACTGACGATGATGACTGATCTATCTCTCAACAC----- 1527
Db      384 ---LeuGluGluLeuGlyThrLeuGlnValThrSerGlnLeuLeuGlnLysAsnThrAsp 402
Qy      1528 GCGCGACGCGCAATGAAGTTTCTA-----AGCTCAGGTAAAGACAGAGCCAAAG 1578
Db      403 ValValAlaThrLeuLysLysLysLeuArgArgTyrlLysAlaLeuLysAspValMetGluLys 422
Qy      1579 CCTAAGAAAGATGAAGATGAAGCAGAGAGCCGAGTCTTCGAAATGGGTGCTCAG 1638
Db      423 AlaAlaGluValTyThrArgLeuLysSerArgValLeuGlyProLysLysLeuAlaVal 442
Qy      1639 CGAGTATTAATAATCTCTTCTGTGCACAGAGACAGCTCCAGAAAAAAGAGACCACA 1698
Db      443 GlnLysValAlaLeuLysAlaGlyMetGluLysGluLysAlaGluGluLysLeu----- 459
Qy      1639 CTGAAGAGGCGAGTGTGTCTCTCGCGGAGTGAAGCACTCGGAAAGGAGCAGCTTGT 1758
Db      460 -----AlaGlyGluGluLeuAlaGlyGluGluAlaProGln 471
Qy      1759 GAGACGACGACGCGCTCGTGGCGGAGCGAT 1788
Db      472 GluLysAlaGluAspLysProSerThrAsp 481

RESULT 12
US-10-203-708-28
; Sequence 28, Application US/10203708
; Publication No. US20030149238A1
; GENERAL INFORMATION:
; APPLICANT: SMITHKLINE BEECHAM CORPORATION
; APPLICANT: SMITHKLINE BEECHAM P.L.C.
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50013
; CURRENT APPLICATION NUMBER: US/10/203,708
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: PCT/US01/04703
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/182,172
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: 60/186,084
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 28
; LENGTH: 676
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-203-708-28

Alignment Scores:      0.00313      Length:      676
Pred. No.:            171.00      Matches:     117
Score:

```

Percent Similarity: 34.32% Conservative: 80  
 Best Local Similarity: 20.38% Mismatches: 215  
 Query Match: 3.65% Indels: 162  
 DB: 14 Gaps: 23

US-09-787-016A-1 (1-2610) x US-10-203-708-28 (1-676)

QY 235 AGTGTCACAGCTTTCAGTGTGGAGCAGATATGGACCAAAAGGCGACCCGAGCAATGAG 294  
 DB 130 AlaValThrAlaThrAlaAlaSerAspArgMetGluSerAspSerAspLysSer 149  
 QY 295 GAGGCACCTAAGGCCATCAAAACCCAGCAGAAAGAGTTCAGGAAAACATGGGGTTTTCGA 354  
 DB 150 SerAspAsnSerGlyLeuLys 162  
 QY 355 AGGACCACTATCCCAAGCAGGAGGCGGAGGAGCGGAGGAGCTGACCCACTGGAGCGG 414  
 DB 163 LysMetSerValSerLysArgAlaAlaArgLysAlaSerAspLysAlaSerVal 182  
 QY 415 CCACCCCCACAGCAGCAGCTGGCGCTGCTCCGCGCAGGTGGAGGAGCCCAAGCGC 474  
 DB 183 SerProSerGluGluGlnSerGluSerSerGluSerGluSerThrSerAspGln 202  
 QY 475 ACTGAGCGGTGAGCAGCTTCCTGACCAT---GCGCGCGCGCGCGC----- 519  
 DB 203 AspPheThrProGluLysLysAlaAlaValAlaArgAlaProArgArgGlyProLeuGlyGly 222  
 QY 520 ---AGAGGAGCATGCTGCTCTCCCTGGAGGATTCGTTGAGCCACGCTCTGC----- 570  
 DB 223 ArgLysLysLysAlaProSerAlaSerAspSerLysAlaSerAspGly 242  
 QY 571 -----CCGCCACAGCGCGAGAGCTCCGAGGCGCGCTGAGGAGCGCT 618  
 DB 243 AlaLysProGluProValAlaMetAlaArgSerAlaSerSerSerSerSerSer 262  
 QY 619 TCT-----GAGACAGAGCGCGCGCGCTGCTCTCCACAGCTGTGAAG 663  
 DB 263 SerSerAspSerValSerValLysLysProProArgLysArgLysProAlaGluLys 282  
 QY 664 -----GACGACCGAGCTCTCTCTGAAAAG 687  
 DB 283 ProLeuProLysProArgLysLysProLysProGluArgProSerSer----- 300  
 QY 688 GTGAAAGGAGGGATGACACGATGACCTCCGATAGTACAGCAGCTGCGCTGCGGAGAG 744  
 DB 301 -----SerSerSerAspSerAspSerAspGluValAspArg 312  
 QY 745 TTGAAAGAGCTTCAGAAATCCG-----CTTCGACGAGAG 777  
 DB 313 IleSerGluTrpLysArgArgAspGluAlaArgArgGluLeuGluAlaArgArg 332  
 QY 778 CGGGAACAGAGCCCACTGAGAGCGCCCTGAAAGGATCCAGATCGCTGCGGAGAGAG 837  
 DB 333 ArgGluGlnGluGluGluLeuArgArgLeuArgGluGlnGluLysGluArg 352  
 QY 838 CGCGCGGAG-----GAGGCTCCGCGCAG-----ACTGCGGCTCCGAGGCCAGT 882  
 DB 353 ArgArgGluAlaAlaSerArgGlyGluAlaGluArgGlySerGlySerGlyAsp 372  
 QY 883 GACACTGTGGAGCGCTCTGCGCAGTAACAGAGCCCGAGAACGATCAGCGGGTGTG 942  
 DB 373 GluLeuArgLysAspAspGluProValLysLysArgGlyArgLysGlyArgGly 392  
 QY 943 TCCGAGCTGGGAAGAT---GACAGAGAGTAAAGTGGAGGAGGAGCGGCTCAGGAC 999  
 DB 393 ProProSerSerSerAspGluProGluAlaGluLeuArgGluAlaLysLysSer 412  
 QY 1000 ATCAAGATGAGGAGCCTGGAGAGCTTGGGCGCAGCGCAAGCTGAATGTGAGGTTACGAC 1059  
 DB 413 AlaLys-----LysProGlnSerSerThrGlu 422  
 QY 1060 CCCAACCCCTGTATTTCATTGCGCCAGCCTCACACACAGGTTATGATTGCTGT 1119

423 Pro-----AlaArgLysProGly----- 428  
 1120 GACCGCTGTGAAGAAATGTTTCATGGCGCATTTGTGTGGCAATTTCTGAGGCTCGAGGGAGG 1179  
 428 ----- 428  
 1180 CTTTGTGAAGAAATGAGGAAAGACTATATCTGCCAAACTGCACCATTTCTGCAAGTGCAG 1239  
 429 -----Gln 429  
 1240 GATGAGACTCANTTCAGAAACCGCGAGATCAGCAGAGAACTAAATGGAGACCTGGAGATGCT 1299  
 430 LysGluLysArgValArgProGluGluLysGlnAlaLys-----ProValLysVal 447  
 1300 GATGCCACCGATTTGTACAGTATAGGAACAATAGAGCAGAAAGTCTAGCGAAGACCAAGCG 1359  
 448 GluArgThr-----ArgLysArgSerGluGlyPheSer 458  
 1360 ATAAAGGTAGAAATGAGAAAGCTGCAAAATCAAGTGGCAAGAGAACTCAAG----- 1413  
 459 MetAspArgLysValGluLysLysLysGluProSerValGluGluLysLysLysLeu 478  
 1414 -----ATCTTCACGCTGTGATAGAGGCGCTGTGTGCTCAAAATGTATTGCG 1461  
 479 HisSerGluLysPheAlaLeuLysValAspSerProAspValLysArgCysLeuAsn 498  
 1462 CCGCGGTCTGTCACTGCGCGCGCGCGCTGCGGTACTGCGAGTAATGACTGTATCCTC 1521  
 499 Ala-----LeuGluGluLeuGlyThrLeuGlnValThrSerGlnLeuGln 514  
 1522 AAACAC-----GCCGCGAGCGCAATGAAGTTCTTA-----AGCTCAGTAAAGAA 1566  
 515 LysAsnThrAspValValAlaThrLeuLysLysLysLysLysLysLysLysLysLysLys 534  
 1567 CAGAGCCCAAGCCCTAAAGAAAGATGAAGATGAAGCCAGCAGAGAGCCCGCTTCCGAAA 1626  
 535 ValMetGluLysAlaAlaGluValTyrThrArgLeuLysSerArgValLeuGlyProLys 554  
 1627 TCGGTGTCTAGCGAGGTATTAAATCTTCTGTGACAGAGAGCCAGCTCCAGAAA 1686  
 555 IleGluAlaValGlnLysValAsnLysAlaGlyMetGluLysGluLysAlaGluGluLys 574  
 1687 AAAGAGACACAGTGAAGAGGCGAGTGTGTCTCGCGCGAGTGAAGTGAAGTGAAGTGAAG 1746  
 575 Leu-----AlaGlyGluGluLeuAlaGlyGlu 583  
 1747 GAAGCAGCTTGTGAGAGCAGCAGCGCTGCTGCGGCGAGCGAT 1788  
 584 GluLeuAlaGlyGluGluAlaProGlnGluLysAlaGluAsp 597

RESULT 13  
 US-09-823-187-87  
 ; Sequence 87, Application US/09823187  
 ; Publication No. US20030096952A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Burgess, Catherine  
 ; APPLICANT: Gusev, Vladimir Y  
 ; APPLICANT: Liu, Xiaohong  
 ; APPLICANT: Majumder, Kumud  
 ; APPLICANT: Padigaru, Muralidhar  
 ; APPLICANT: Patturajan, Meera  
 ; APPLICANT: Shinkets, Richard A  
 ; APPLICANT: Spaderna, Steven K  
 ; APPLICANT: Spyttek, Kimberly  
 ; APPLICANT: Taupier, Raymond J  
 ; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME  
 ; FILE REFERENCE: 15966-745  
 ; CURRENT APPLICATION NUMBER: US/09/823,187  
 ; CURRENT FILING DATE: 2001-03-29  
 ; PRIOR APPLICATION NUMBER: 60/193,339  
 ; PRIOR FILING DATE: 2000-03-30  
 ; PRIOR APPLICATION NUMBER: 60/193,205  
 ; PRIOR FILING DATE: 2000-03-30



; PRIOR APPLICATION NUMBER: 60/195,343  
 ; PRIOR FILING DATE: 2000-04-05  
 ; PRIOR APPLICATION NUMBER: 60/195,088  
 ; PRIOR FILING DATE: 2000-04-06  
 ; PRIOR APPLICATION NUMBER: 60/195,005  
 ; PRIOR FILING DATE: 2000-04-06  
 ; PRIOR APPLICATION NUMBER: 60/195,792  
 ; PRIOR FILING DATE: 2000-04-10  
 ; PRIOR APPLICATION NUMBER: 60/196,556  
 ; PRIOR FILING DATE: 2000-04-11  
 ; PRIOR APPLICATION NUMBER: 60/197,081  
 ; PRIOR FILING DATE: 2000-04-13  
 ; PRIOR APPLICATION NUMBER: 60/197,525  
 ; PRIOR FILING DATE: 2000-04-14  
 ; PRIOR APPLICATION NUMBER: 60/197,087  
 ; PRIOR FILING DATE: 2000-04-14  
 ; NUMBER OF SEQ ID NOS: 103  
 ; SOFTWARE: Patentin Ver. 2.1  
 ; SEQ ID NO 87  
 ; LENGTH: 678  
 ; TYPE: PRT  
 ; ORGANISM: Mus musculus  
 US-09-823-187-87

Alignment Scores:  
 Pred. No.: 0.00313 Length: 678  
 Score: 171.00 Matches: 135  
 Percent Similarity: 32.25% Conservative: 83  
 Best Local Similarity: 19.97% Mismatches: 194  
 Query Match: 3.65% Indels: 264  
 DB: 10 Gaps: 32

US-09-787-016a-1 (1-2610) x US-09-823-187-87 (1-678)

QY 214 GRATTAGGATTTCAGGAAAGTGTCCAGCTTTCAGTGTGGAGCAGGTATGAGCAGC 273  
 Db 125 ValMetThrValThrAlaValThrThrAlaThrSerAspArgMetGluSerAspSer 144  
 QY 274 AAAGGCGACCCGAGCAATGAGGAGGACCTAAGGCCATCAAAACCCAGCAAGAGTTC 333  
 Db 145 AspSerAspSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 153  
 QY 334 AGGAAACATGGGGTTTTCGAGGACCACT-ATGCCCAAGCGA 375  
 Db 154 -----GlyLeuLysArgLysThrProValLeuLysValSerValSerLysArg 169  
 QY 376 GAGGCGCGAGGACCGGAGGCTGACCCACTGGAGCCGCCACCCACAGCAGCAGCTG 435  
 Db 170 AlaArgAlaSerSerAspLeuAspGlnAlaSerValSerProSerGluGlu----- 187  
 QY 436 GGCCTGTCCCTGCGCGCAGTGGGAGGAGCCAGCCAGCGCACTGAGCGCGTGGAG----- 489  
 Db 188 -----AspSerGluSerProSerGluSerGluSerLysThrSerAspGln 201  
 QY 490 -----CAGTTCTGACCATTCGCGG-----CGCGCGGAGGAGGAGCATG 531  
 Db 202 AspPheThrProGluLysLysThrAlaAlaArgProArgArgGly----- 217  
 QY 532 CCGTCTCCTCGAGGATTCTGTGTGAGCCACG-----TCCTGCCCGCGC----- 576  
 Db 218 ProLeuGlyArgLysLysLysHisProThrGlyTyraLysProGluLysValPro 237  
 QY 577 -----ACAGACCGCGAGCAGCGCTCCGAGCGAGCGTGGAA----- 612  
 Db 238 SerAlaSerAspSerSerLysAlaAspSerAspGlyAlaLysGluGluProValVal 257  
 QY 613 -----AGCGCTTCTGAGCAGGAGCGCGCCCGCCAGTCTGCT 648  
 Db 258 ThrAlaGlnProSerProSerSerSerSerSerSerSerSerSerSerSerSerSer 277  
 QY 649 TCCACAGCTGTGAAG----- 663  
 Db 278 AspValSerValLysLysProProArgGlyArgLysProAlaGluLysProProLys 297

QY 664 -----GAACGACACAGCCTCTCTGAAAAGGTGAAGAGGG 699  
 Db 298 ProArgGlyArgProLysProGluArgProSerThr----- 311  
 QY 700 GATGACACGATGACACCTCCGATAGTCACAGCATGCGCTGACCTGAAGAGCTTCAG 759  
 Db 312 -----SerSerSerAspSerAspSerGlyGluValAspArgLys 327  
 QY 760 AAT-----CGCTTCGAGGAGCGGAA 783  
 Db 328 GluTrpLysArgAspGluGluArgArgGluLeuGluAlaArgArgArgGlu 347  
 QY 784 CAGAGCCCACTGAGAGCCCTGAAAGGATCCAGAGTCGCTCGGAGAGCGCGG 843  
 Db 348 GlnGluGluLeuArgLysGluGluGln-----GlnGluArgGluGluGluArg 366  
 QY 844 GAGAGGCTCCCGCGAGACTGTGGCTCCGAGGCCAGTGCACCTGTGGAGGCGCTCTG 903  
 Db 367 LysGlu-----ArgAlaGluArgGlySer----- 375  
 QY 904 CCCAGTAGCAGGAGCCCGAGAACGATCAGGGGTTGTGCCAGGCTGGGAAAGATGAC 963  
 Db 376 ---SerGlyGluGluLeuAspGluProVal-----LysLys 388  
 QY 964 AGAGAGAGTAAAGTTGAGGCGAAAGCGCTCAGGACATCAAAGATGAGAGCCT---GGA 1020  
 Db 389 ArgSerArgLysAlaArgGlyArgGlyThrProSerSerSerAspSerGluProGluGly 408  
 QY 1021 GACTTGGGCCGA-----CCGAAAGCTGAATGTGAGGCTTACAGCCCC 1062  
 Db 409 GluLeuGlyLysGluGlyLysLysLysLysLysSerGlnLeuProGlySerGluSer 428  
 QY 1063 AACGCCCTGATTGCTATTTGCGCGCAGCTCACAACACAGGTTTATGATTGCTGTGAC 1122  
 Db 429 -----AlaArgLysPro----- 432  
 QY 1123 CGCTGTGAAGAATGTTTCATGGCGATTGTGTGGGCATTTCTGAGGCTCAGGAGGCTT 1182  
 Db 433 -----GlyGlnLysGluLysArgGlyArg--- 440  
 QY 1183 TTGGAAGGAATGGGAAGACTATATCTGCCAAACTGCACCATTTCTGCAAGTGCAGGAT 1242  
 Db 440 ----- 440  
 QY 1243 GAGACTCATTGAGAAACCGCAGATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTAT 1302  
 Db 441 -----ProAspGluLysProArgAlaArgProValLysValGlu 453  
 QY 1303 GGCACCGATGTACAGTATAGGAACAATAGCAGAGAGTCTAGCAGAGACCAAGGATA 1362  
 Db 454 ArgThr-----ArgLysArgSerGluGlyLeuSerLeu 464  
 QY 1363 AAGGCTAGATTTGAGAAAGCTCCTAAATCCAGTGGCAAGAGAACTCAAG----- 1413  
 Db 465 GluArgLysGlyGluLysLysLysGluProSerValGluArgLysGluLysLeuHis 484  
 QY 1414 -----ATCTTCAGCCTGTGTAGAGCGCTGTGCTCAAAATGTATTGTATGCC 1464  
 Db 485 SerGluIleLysPheAlaLeuLysValAspAsnProAspValArgLysCysLeuSerAla 504  
 QY 1465 -----GGTGTGTGTACCTGGCG-----CAGCCCGACTCG 1494  
 Db 505 LeuGluGluLeuGlyThrLeuGlnValThrSerGlnIleLeuGlnLysAsnThrAspVal 524  
 QY 1495 GTG-----TACTCAGTAATGACTGTATCTCAACACCGCC 1530  
 Db 525 ValAlaThrLeuLysLysIleArgArgTyrLysAlaAsnLysAspValMetAlaLysAla 544  
 QY 1531 GCA-----CGCAATG 1542  
 Db 545 AlaGluValTyrThrArgLeuLysSerArgValLeuGlyProLysValGluAlaLeuGln 564

```
QY 1543 AAGTTCTTAAGCTCAGTAAAGACAGCAAGCCAAAGCCTAAAGAAAGATGAAGATGAAG 1602
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 565 LysValAsnLysAlaGlyAlaGluLysGluArgAlaAspAsnGluLysLeuGluGln 584
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1603 CAGAGAACCCAGTCTTCGAAATCGCGTGCAGCAGGATTTAAATCTCTCTCTG 1662
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 585 Pro-----GlyGluAlaProArgGluLeuAlaGluAsp 596
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1663 CAC-----AAGACAGCAGCTCCAGAAAGAGAGACACAGTGAAGAGGAGCG 1713
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 597 GluProSerThrAspSerAlaProValAsnGlyGluAlaThrSerGlnLysGlyGlu 616
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1714 GTGTCCTCGCGGAGTGAAGCTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1773
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 617 AspMetGluAspArgAlaGlnGluAspGlyGlnAspSer-----GluAspGlyProArg 634
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1774 TCGTGGCGAGCAGTAC-----AAATACAAATGACGTAAGACCA 1812
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 635 GlyGlySerSerGluLeuHisAspSerProArgAspAsnSerAspProAlaLysPro 654
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1813 -----GAAAGACTGCTGCTCCCTCGCCGCTCA 1839
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 655 GlyAsnGluArgGlnAspHisGluArgThrArgLeuAlaSerGluSer 670
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
US-09-863-776-52
; Sequence 52, Application US/09863776
; Publication No. US20030198953A1
; GENERAL INFORMATION:
; APPLICANT: Spytex, Kimberly A
; APPLICANT: Majumder, Kumud
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Mishra, Vishnu
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Spaderna, Steven K
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Rastelli, Luca
; APPLICANT: Li, Li
; APPLICANT: Taupier, Raymond J
; APPLICANT: Gangoli, Esha
; TITLE OF INVENTION: No. US20030198953A1el Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-020
; CURRENT APPLICATION NUMBER: US/09/863, 776
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 09/540, 763
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 60/206, 679
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: 60/206, 688
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: 60/206, 829
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: 60/207, 748
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: 60/207, 798
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: 60/208, 263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 60/208, 831
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: 60/209, 451
; PRIOR FILING DATE: 2000-06-05
; PRIOR APPLICATION NUMBER: 60/210, 060
; PRIOR FILING DATE: 2000-06-07
; PRIOR APPLICATION NUMBER: 60/219, 507
; PRIOR FILING DATE: 2000-07-20
; PRIOR APPLICATION NUMBER: 60/221, 337
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/221, 927
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 60/263, 135
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/263, 688
```

```
; PRIOR FILING DATE: 2001-01-24
; PRIOR APPLICATION NUMBER: 60/263, 694
; PRIOR FILING DATE: 2001-01-24
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 52
; LENGTH: 678
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-863-776-52

Alignment Scores:
Pred. No.: 0.00313 Length: 678
Score: 171.00 Matches: 135
Percent Similarity: 32.25% Conservative: 83
Best Local Similarity: 19.97% Mismatches: 194
Query Match: 3.65% Indels: 264
DB: 10 Gaps: 32

US-09-787-016A-1 (1-2510) x US-09-863-776-52 (1-678)
QY 214 GTATTAGATTTCAGGGAAGAGTGTCCAGCTTTCAGTTTGGAGCAGGTATGGACGAC 273
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125 ValMetThrValThrAlaValThrThrAlaThrSerAspArgMetGluSerAspSer 144
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 274 AAAGCGCAGCCGACCAATGAGGAGGACCTAAGGCCATCAAAACCACCCAGCAAGAGTTT 333
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 145 AspSerAspLysSerSerAspHisSer----- 153
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 334 AGGAAACATGGGGTTTCGAGAGGACCACT-----ATGCCCAAGCGA 375
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 154 -----GlyLeuLysArgLysThrProValLeuLysValSerValSerLysArg 169
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 376 GAGGCGGAGGCGGAGGAGCTGACCCACTGAGAGCGGCCACCCACCCAGCAGCAGCGTG 435
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 170 AlaArgArgAlaSerSerAspLeuaspGlnAlaSerValSerProSerGluGlu 187
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 436 GGCCTGTCCCTGCGCGGAGTGGGAGGCGGCCAGCCAGCCAGCAGCGCGCTGCAG----- 489
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 188 -----AspSerGluSerProSerGluSerGluLysThrSerAspGln 201
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 490 -----CACTTCTGACCATTTGCGCG-----CGCGCGGCGGAGGAGGAGCGATG 531
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 202 AspPheThrProGluLysLysThrAlaAlaArgProProArgArgGly----- 217
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 532 CCTGTCTCCCTGAGGATTTGTTGAGCCGACG-----TCTGCCCCGCC----- 576
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 218 ProLeuGlyGlyArgLysLysLysHisProThrGlyfYrAlaCysProGlnLysValPro 237
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 577 -----ACAGAGCCGAGACAGCTCCGAGGCGAGCGTGGAA----- 612
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 238 SerAlaSerAspSerAspSerLysAlaAspSerAspGlyAlaLysGluProValVal 257
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 613 -----AGCGCTTCTGAGACCAAGAGCGGCCCCCAGTCTGCT 648
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 258 ThrAlaGlnProSerProSerSerSerSerSerSerSerSerSerSerSerSer 277
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 649 TCACAGCTGTGAAG----- 663
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 278 AspValSerValLysLysProProArgGlyArgLysProAlaGluLysProProLys 297
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 664 -----GAAAGCAGGAGCTCTTCTGAAAGAGGTGAAGAGAGG 699
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 298 ProArgGlyArgArgProLysProLysProLysProLysProLysProLysProLys 311
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 700 GATGACACGATGACACTCCGATAGTACACGATGGCTGACCTTGAAGAGCTTCAG 759
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 312 -----SerSerSerAspSerAspSerGlyGluValAspArgLysSer 327
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 760 AAT-----CGCCTTCGAGGAGCGGGA 783
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 328 GluTrpLysArgArgAspGluGluArgArgGluLeuGluAlaArgArgArgGlu 347
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

QY 784 CAGGAGCCCACTGAGAGGCCCTGAAAGGGATCCAGAGTCGCTGGGAAAGAGCCGCG 843  
Db 348 GlnGluGluGluLeuArgArgLeuArgGlu--GlnGluArgGluGluGluArgGluArg 366  
QY 844 GAGGAGGGTCCCGCCGAGACTGTGGCTCCGAGGCCAGTGCACAGTGTGGAGGCGCTCTG 903  
Db 367 LysGlu--ArgAlaGluArgGlyGlySer 375  
QY 904 CCAGTAAGCAGAGAGCCGAGAACGATCAGAGGGGTTGTCTCCAGCTGGGAAAGATGAC 963  
Db 376 ---SerGlyGluGluLeuGluArgGluGluProVal--LysLys 388  
QY 964 AGAGAGAGTAAGTTGAGAGGAAGCGGCTCAGGACATCAAGATCAGAGAGCT--GGA 1020  
Db 389 ArgSerArgLysAlaArgGlyArgGlyThrProSerSerSerSerGluProGluGly 408  
QY 1021 GACTTCGGCGGA 1062  
Db 409 GluLeuGlyLysGluGlyLysLysLeuAlaLysLysSerGluLeuProGlySerGluSer 428  
QY 1063 AACGCCCTGTATTGCTATTCGCGCCAGCTCACAACACAGGTTTATGTTGCTGTGAC 1122  
Db 429 -----AlaArgLysPro----- 432  
QY 1123 CGCTGTGAAGATGTTTCATGCGGATTTGTGGGCAATTTCTGAGGCTCGAGGAGGCTT 1182  
Db 433 -----GlyGlnLysGluLysArgGlyArg--- 440  
QY 1183 TTGGAAGGAATGGGAAGACTATATCTGCCAAACTGCACCAATTCGCAAGTGCAGGAT 1242  
Db 440 ----- 440  
QY 1243 GAGACTCATTCAAGACGGCAGATCAGCAGAGCTAATGAGACCTCGAGATGCTGAT 1302  
Db 441 -----ProAspGluLysProArgAlaArgProValLysValGlu 453  
QY 1303 GGCACCGATTGTACAAATATAGGAACAATAGACGAGAAGTCTAGCGAAGACCAAGGATA 1362  
Db 454 ArgThr-----ArgLysArgSerGluGlyLeuSerLeu 464  
QY 1363 AAGGCTAGATTCAGAAAGCTCAATCCAGTGGCAGAGAACTCAAG----- 1413  
Db 465 GluArgLysGlyGluLysLysLysGluProSerValGluGluArgGluLysLeuHis 484  
QY 1414 -----ATCTTCGAGCCTGTGATAGAGCGCCTGGTCCCTCAAAATGTATGCGCC 1464  
Db 485 SerGluLysPheAlaLeuLysValAspAsnProAspValArgLysCysLeuSerAla 504  
QY 1465 -----CGCTGCTGTCCGTCGCG-----CAGCCCGACTCG 1494  
Db 505 LeuGluGluLeuGlyThrLeuGluValThrSerGlnLeuGlnLysAsnThrAspVal 524  
QY 1495 GTG-----TACTGCAGTAATGACTGTATCTCCAAACACGCG 1530  
Db 525 ValAlaThrLeuLysLysIleArgArgLysAlaAsnLysAspValMetAlaLysAla 544  
QY 1531 GCA-----GCGACAATG 1542  
Db 545 AlaGluValThrArgLeuLysSerArgValLeuGlyProLysValGluAlaLeuGln 564  
QY 1543 AAGTTCTAAGCTCAGTAAGACAGACCAAGCCTAAGCAAAAGATGAGATGAG 1602  
Db 565 LysValAsnLysAlaGlyAlaGluLysGluArgAlaAspAsnGluLysLeuGluGln 584  
QY 1603 CCAGAGAAGCCAGTCTTCGAAATCGGTGCTCAGGCGAGTATTAATCTCTCTGTG 1662  
Db 585 Pro-----GlyGluGlnAlaProArgGluLeuAlaGluAsp 596  
QY 1663 CAC-----AAGACACAGCTCCGAAAAAAGACACCAAGTGAAGAGGAGTGTG 1713  
Db 597 GluProSerThrAspArgSerAlaProValAsnGlyGluAlaThrSerGlnLysGlyGlu 616  
QY 1714 GTGGTCCCTCGCGGAGTGAAGCACTCGGGAAGGAAGACAGCTTGTGAGCAGCAGCGCG 1773

Db 617 AsnMetGluAspArgAlaGlnGluAspGlyGlnAspSer-----GluAspGlyProArg 634  
QY 1774 TCGTGGCGCAGCGATCAC-----AATTCAATGCGAGTAAAGCCA 1812  
Db 635 GlyGlySerSerGluGluLeuHisAspSerProArgAspAsnSerAspProAlaLysPro 654  
QY 1813 -----GAAAAGACTGCTGCTCCCTCGCGCTCA 1839  
Db 655 GlyAsnGluArgGlnAspHisGluArgThrArgLeuAlaSerGluSer 670  
RESULT 15  
US-10-203-708-29  
; Sequence 29, Application US/10203708  
; Publication No. US20030149238A1  
; GENERAL INFORMATION:  
; APPLICANT: SMITHKLINE BEECHAM CORPORATION  
; TITLE OF INVENTION: NOVEL COMPOUNDS  
; FILE REFERENCE: GP50013  
; CURRENT APPLICATION NUMBER: US/10/203,708  
; CURRENT FILING DATE: 2002-08-13  
; PRIOR APPLICATION NUMBER: PCT/US01/04703  
; PRIOR FILING DATE: 2001-02-14  
; PRIOR APPLICATION NUMBER: 60/182,172  
; PRIOR FILING DATE: 2000-02-14  
; PRIOR APPLICATION NUMBER: 60/186,084  
; PRIOR FILING DATE: 2000-02-29  
; NUMBER OF SEQ ID NOS: 46  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 29  
; LENGTH: 717  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-203-708-29  
Alignment Scores:  
Pred. No.: 0.00317 Length: 717  
Score: 171.00 Matches: 117  
Percent Similarity: 34.32% Conservative: 80  
Best Local Similarity: 20.38% Mismatches: 215  
Query Match: 3.65% Indels: 162  
DB: 14 Gaps: 23  
US-09-787-016A-1 (1-2610) x US-10-203-708-29 (1-717)  
QY 235 AGTGTCCCAAGCTTTCAGTGTGTGGAGCAGTATGACGACAAAGGCGACCCGAGCAATGAG 294  
Db 171 AlaValThrAlaThrAlaAlaSerAspArgMetGluSerAspSerAspSerAspLysSer 190  
QY 295 GAGGCACCTTAGGCGCATCAACCCACCAGCAAGAGATTGAGAAACATGGGGTTTCCA 354  
Db 191 SerAspAsnSerGlyLeuLys-----ArgLysThrProAlaLeu--- 203  
QY 355 AGGACCCTATCGCAAGCGAGAGCGCGCGGAGCGGCGGAGCTGACCCACTGGAGCGG 414  
Db 204 LysMetSerValSerLysArgAlaArgLysAlaSerSerAspLeuAspGlnAlaSerVal 223  
QY 415 CCACCCCGCCACAGCAGCTGGGCTGCTCCCTGCGCGCAGTGGGAGGCGCCAGCCGCGC 474  
Db 224 SerProSerGluGluAsnSerGluSerSerGluSerGluLysThrSerAspGln 243  
QY 475 ACTGAGCGGTGAGCGTCTCTGACCATTT---GGCGCGCGCGCGCG--- 519  
Db 244 AspPheThrProGlnLysLysAlaAlaValArgAlaProArgArgGlyProLeuGlyGly 263  
QY 520 ---AGGAGGAGCATGCTGTCTCTCGGAGGATTCGTGTGAGCCCGCCAGCTCTCTGCG--- 570  
Db 264 ArgLysLysLysLysAlaProSerAlaSerAspSerLysAlaAspSerAspGly 283  
QY 571 -----CCGCCACAGACCGGAGCAGCTCCGAGGCGAGCGCTGGAACGCGCT 618  
Db 284 AlaLysProGluProValAlaMetAlaArgSerAlaSerSerSerSerSerSerSer 303

QY 619 TCT-----GAGACCAAGAGCGCCCGCCAGTCTGTCTCCACAGCTGTGAAG 663  
Db 304 SerSerAspSerAspValSerValLysLysProProArgGlyArgLysProAlaGluLys 323  
QY 664 -----GACGACCAAGCTCTTCTTCTGAAGA 687  
Db 324 ProLeuProLysProArgGlyArgLysProLysProGluArgProProSerSer 341  
QY 688 GTCAAGAGGAGGGATGACCAAGATGACCTCCGATAGTACAGAGTGGCCCTG--ACC 744  
Db 342 -----SerSerSerAspSerAspSerAspGluValAspArg 353  
QY 745 TTCAAGAGCTTCCAGATCCG-----CTTCGCGGGAAG 777  
Db 354 IleSerGluTriLysArgArgAspGluAlaArgArgGluLeuGluAlaArgArgArg 373  
QY 778 CGGACACAGAGCCCACTGAGGCGCCCTGAAGAGGATCCAGAGTGGCTCGGGAAGAG 837  
Db 374 ArgGluGlnGluGluLeuArgArgLeuArgGluGlnGluLysGluGluArg 393  
QY 838 CGCGCGGAG-----GAGGCTCCCGCGAG-----ACTGTGGGCTCCGAGGCCAGT 882  
Db 394 ArgArgGluArgAlaAspArgGlyGluAlaGluArgGlySerGlyGlySerGlyAsp 413  
QY 883 GACACTGTGGAGGCGTCTGCCAGTATAGCAGGAGCCCGAGNACATCAGGGGTTGTG 942  
Db 414 GluLeuArgGluAspAspGluProValLysLysArgGlyArgLysGlyArgGly 433  
QY 943 TCCAGCGTGGGAAGAT---GACAGAGAGATGAGTGGAGGGAAGCGGCTCAGGAC 999  
Db 434 ProProSerSerSerAspSerGluProGluAlaGluLeuGluArgGluAlaLysLysSer 453  
QY 1000 ATCAAGATGAGGAGCGTGGAGCTTGGCGCGGAGCCGAGCTGAATGTGAGGGTTACGAC 1059  
Db 454 AlaLys-----LysProGlnSerSerSerThrGlu 463  
QY 1060 CCCAAGCCCTGTATTGCAATTTGCCCGCAGCTCACACACAGGTTTATGATTGCTGT 1119  
Db 464 Pro-----AlaArgLysProGly----- 469  
QY 1120 GACCGCTGTGAAGATGGTTTCATGCGCATTTGTGGGCATTTCTGAGGCTGAGGGAGG 1179  
Db 469 ----- 469  
QY 1180 CTTTTCGAAGGANTGGGAGACTATATCTGCCAAACCTGCACCATTTCTGCAAGTGACG 1239  
Db 470 -----Gln 470  
QY 1240 GATGAGACTCATTCAAGAAACGGCAGATCAGCAGAAAGCTAAATGGAGACTGGAGATGCT 1299  
Db 471 LysGluLysArgValArgProGluGluLysGlnAlaLys-----ProValLysVal 488  
QY 1300 GATGGCCCGCATTTGTACAGTATAGGACAAATAGACGACAGAGTCTAGCGAGACCAAGG 1359  
Db 489 GluArgThr-----ArgLysArgSerGluGlyPheSer 499  
QY 1360 ATAAAGGGTAGAATTGAGAAAGCTGCAAAATCCAAGTGGCAAGAGAACTCAAG----- 1413  
Db 500 MetAspArgLysValGluLysLysLysLysGluProSerValGluGluLysLeuGlnLysLeu 519  
QY 1414 -----ATCTTCCAGCCTGTGATAGAGCGCTCGTGCCTCAAAATGATTGGC 1461  
Db 520 HisSerGluIleLysPheAlaLeuLysValAspSerProAspValLysArgCysLeuAsn 539  
QY 1462 CCGGGTGCTGTACGTGGCGCGAGCCCGAGTCTGGTGACTGACATGATGATGATCTCTC 1521  
Db 540 Ala-----LeuGluGluLeuGlyThrLeuGlnValThrSerGlnIleLeuGln 555  
QY 1522 AAACAC-----GCCGAGCGACAAATCAAGTTCTTA-----AGCTCAGGTAAAGAA 1566  
Db 556 LysAsnThrAspValValAlaThrLeuLysLysIleArgGlyrLysAlaAsnLysAsp 575

QY 1567 CAGAAGCCAAAGCCTAAAGAAAGATGAAGTGAAGCCAGAGCCAGTCTTCCGAAA 1626  
Db 576 ValMetGluLysAlaAlaGluValThrArgLeuLysSerArgValLeuGlyProLys 595  
QY 1627 TCGGTGCTCAGGCGAGGTATTAAATCTCTTCTGTGCACAGAGACCCAGCTCCAGAAAA 1686  
Db 596 IleGluAlaValGlnLysValAsnLysAlaGlyMetGluLysGluLysAlaGluGluLys 615  
QY 1687 AAAGAGACCAACAGTGAAGAGGCGAGTGTGTCTCCCTGGCGGAGTGAAGCACTCGGAAG 1746  
Db 616 Leu-----AlaGlyGluGluLeuAlaGlyGlu 624  
QY 1747 GAAGCAGCTTGTGAGAGCAGCACGCCGCTGCTGGCGGAGCGAT 1788  
Db 625 GluLeuAlaGlyGluGluAlaProGlnGluLysAlaGluAsp 638

Search completed: April 28, 2004, 11:22:28  
Job time : 242.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 28, 2004, 10:49:27 ; Search time 29.5 Seconds  
(without alignments)  
9135.172 Million cell updates/sec

Title: US-09-787-016A-1  
Perfect score: 4689  
Sequence: 1 ctcggggcgctcgccacc.....actcttaagatcatatctg 2610

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delc 6.0, Delcxt 7.0

Searched: 389414 seqs, 51625971 residues  
Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool\_p/US09787016/runat\_28042004\_094330\_26973/app\_query.fasta\_1.2759  
-DB=Issued Patents AA -OPMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=100 -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09787016 -CGN 1 1 29 -runat 28042004\_094330\_26973 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A.COMB.pcp.\*  
2: /cgn2\_6/ptodata/2/iaa/5B.COMB.pcp.\*  
3: /cgn2\_6/ptodata/2/iaa/6A.COMB.pcp.\*  
4: /cgn2\_6/ptodata/2/iaa/6B.COMB.pcp.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS.COMB.pcp.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	174.5	3.7	2289	3 US-09-051-019-2	Sequence 2, Appli
2	171	3.6	700	4 US-09-252-991A-32447	Sequence 32447, A
3	169	3.6	1404	4 US-08-801-308-1	Sequence 1, Appli
4	164.5	3.5	1075	4 US-09-252-991A-18387	Sequence 18387, A
5	164.5	3.5	1427	4 US-09-252-991A-27005	Sequence 27005, A
6	163.5	3.5	1780	1 US-08-769-309A-5	Sequence 5, Appli
7	163.5	3.5	1780	3 US-08-994-570-5	Sequence 5, Appli
8	162.5	3.5	1805	1 US-07-853-913-2	Sequence 2, Appli
9	162	3.5	341	4 US-09-252-991A-22507	Sequence 22507, A
10	161	3.4	238	4 US-09-257-179-32	Sequence 80, Appli
11	160	3.4	8991	4 US-08-714-741-32	Sequence 32, Appli
12	158.5	3.4	779	4 US-10-164-595-56	Sequence 56, Appli

13	156.5	3.3	778	1 US-08-439-818A-5	Sequence 5, Appli
14	156.5	3.3	778	2 US-08-751-965-5	Sequence 5, Appli
15	156.5	3.3	778	2 US-08-738-375-5	Sequence 5, Appli
16	156.5	3.3	778	2 US-08-728-626-5	Sequence 5, Appli
17	156.5	3.3	778	3 US-08-808-599A-5	Sequence 5, Appli
18	154.5	3.3	1079	4 US-09-489-039A-7502	Sequence 7502, Ap
19	154.5	3.3	1618	1 US-07-853-913-4	Sequence 4, Appli
20	153.5	3.3	1596	4 US-08-978-277A-4	Sequence 5, Appli
21	153	3.3	3969	3 US-08-061-376-5	Sequence 5, Appli
22	152.5	3.2	442	4 US-09-252-991A-23285	Sequence 23285, A
23	152.5	3.2	632	4 US-09-252-991A-23129	Sequence 23129, A
24	152.5	3.3	735	4 US-10-164-595-50	Sequence 80, Appli
25	152.5	3.3	784	4 US-10-164-595-79	Sequence 79, Appli
26	152.5	3.3	843	4 US-10-164-595-54	Sequence 54, Appli
27	152.5	3.3	904	4 US-08-976-594-815	Sequence 615, App
28	152.5	3.3	1187	1 US-08-320-559-28	Sequence 28, Appli
29	152.5	3.3	1187	3 US-08-545-860D-28	Sequence 28, Appli
30	152.5	3.3	1187	5 PCT-US94-04496-28	Sequence 26, Appli
31	152.5	3.3	1210	1 US-08-320-559-26	Sequence 26, Appli
32	152.5	3.3	1210	3 US-08-545-860D-26	Sequence 26, Appli
33	152.5	3.3	1210	5 PCT-US94-04496-26	Sequence 26, Appli
34	152	3.2	2468	4 US-09-976-594-726	Sequence 726, App
35	151	3.2	871	4 US-09-252-991A-19431	Sequence 19431, A
36	150	3.1	226	4 US-09-252-991A-29689	Sequence 29689, A
37	147.5	3.1	1051	4 US-09-428-711A-14	Sequence 14, Appli
38	147.5	3.1	1061	4 US-09-762-481B-2	Sequence 2, Appli
39	147	3.1	504	4 US-09-252-991A-28242	Sequence 28242, A
40	146.5	3.1	663	4 US-09-252-991A-30843	Sequence 30843, A
41	146.5	3.1	732	1 US-08-317-522A-5	Sequence 5, Appli
42	146.5	3.1	1002	4 US-09-252-991A-27980	Sequence 27980, A
43	145	3.1	836	4 US-09-252-991A-26065	Sequence 26065, A
44	144	3.0	253	4 US-09-252-991A-31638	Sequence 31638, A
45	144	3.0	469	4 US-09-252-991A-32910	Sequence 32910, A

ALIGNMENTS

RESULT 1  
US-09-051-019-2  
Sequence 2, Application US/09051019  
Patent No. 6103229  
GENERAL INFORMATION:  
APPLICANT: KAHMANN, Regine and QUADBECK-SERGER, Claudia  
TITLE OF INVENTION: Regulatory gene from Ustilago maydis  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESSES:  
ADDRESSES: Keil & Weinkauff  
STREET: 1101 Connecticut Avenue  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.2 Mb storage  
COMPUTER: IBM AT-compatible, Pentium processor  
OPERATING SYSTEM: Windows 98  
SOFTWARE: WordPerfect version 6.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/051,019  
FILING DATE: 31-MAR-1998  
CLASSIFICATION:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2289 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-09-051-019-2  
Alignment Scores:  
Pred. No.: 6.58e-06  
Score: 174.50  
Percent Similarity: 34.28%  
Best Local Similarity: 20.07%  
Length: 2289  
Matches: 120  
Conservative: 85  
Mismatch: 184

Query Match: 3.72% Indels: 209  
DB: 3 Caps: 28

US-09-787-016A-1 (1-2610) x US-09-051-019-2 (1-2289)

QY 346 GGTGTTTCAGAGGACCACTATCGCCAGCAGAGGCGCGAGGAGCCGCGAGGCTGACCCA 405  
DB 1387 GlySerArgAlaLeuThrAspAlaAspArg-AspGluGluLeuAsp- 1401

QY 406 CTGGAGCCGCCACCCACACAGCAGCAGCTGGGCGCTGCTCCCTGGCGCAGTGGGAGGCAG 465  
DB 1401 - 1401

QY 466 CCCAAGCGCACCTGAGCGGTGGAGCAGTTCCTGACCATTCGCGCGCGCGCGCGAGGAGG 525  
DB 1402 --LysLeuValGluArgValGluAspAlaThrAspProAlaAspAspGlnAsnLysPro 1420

QY 526 AGCATGCTGCTCTCCCTGAGGAGTTCGTGAGCCGACGCTGCTGCGCCGCCACAGAGCC 585  
DB 1421 AsnAlaArgAsnCysIleCysArgSerMetProIleAlaIleProSerSerGly 1440

QY 586 GAGACAGCTCCGAGGCGAGCGGAA-AGCGTCTCAGACACAGAGCGGC 636  
DB 1441 AlaGluCysSerArgCysArgValGlnTyrHisLeuSerCysIleLysValArgSerSer 1460

QY 637 CCCAGTCT- 651  
DB 1461 GluValSerArgAlaGluGlyGlyTrpValCysProPheCysProTrpTyrGlySerAla 1480

QY 652 ACAGCTGTGAGGAAGCAGCAGCTCTCT- 699  
DB 1481 ProPheLeuLysMetArgLysAlaIleSerIleAlaAspLeuSerLysLeuValTyrAsp 1500

QY 700 GATGACAC- 753  
DB 1501 GlnAspHisArgAspGlnPheLysPheLeuProLeuGluTrpAspAlaIleGluGlu 1520

QY 754 CTTGAGATCGCTTCGCGAAGCGGAGCAGAGCCGCTCAGAGGCGCCCTGGAAGG 813  
DB 1521 ValVal- 1538

QY 814 ATCCAGAGTCCGCGGAGAGCGCGGAGGAG- 849  
DB 1539 Leu- 1557

QY 850 - 861  
DB 1558 SerIleGlyCysProValAspValLeuGlyProGluLysValAsnMetLeuAspLeuIle 1577

QY 862 - 888  
DB 1578 SerGluAsnLeuLeuAlaLeuGlySerGlnGlnClyAspAlaProMetAlaProVal 1597

QY 889 - 918  
DB 1598 GluArgIleLysAlaSerThrProAlaArgSerAspGluArgThrGluGluThrThrPro 1617

QY 919 - 948  
DB 1618 LeuProArgSerSerArgValProAlaProAlaAspArgSerGly- 1635

QY 949 GCTGCGAAGATGACAGAGAGTAAGTTCGAG- 1005  
DB 1636 AlaValArgAspAspArgLysAlaLysArgGlyLysArgAlaLysLeuValPhe 1655

QY 1006 GATGAGAGCTGAGAGCTGGGCGGCGGAGCCTGAAATGTGAGGCTTACGACCCCAAC 1065  
DB 1656 GlnGluGluIleGly- 1668

QY 1066 GCCCTGTATTCATTTGCGCGCAGCCTCACAACACAGGTTTATGATTTCGTGACCCG 1125  
DB 1669 ProIleTyrCysLeuCysHisGluProGluSerGlyArg- 1687

QY 1126 TGTGAAGAATGTTTCATGGCGGATGTGTGGCCATTCTGAGGCTCGAGGGAGGCTTTTG 1185  
DB 1688 CysMetLeuTrpPheHisThrAsnCysValArgLeuAspAspPro- 1703

QY 1186 GAAAGGAATGGGAAGACTATATCGCCAACTCAACCTGCACCATTTCTCAAGTCGAGGATGAG 1245  
DB 1704 AsnLeuGlyAsnGluProTrpIleCysProMetCysCysIle- 1717

QY 1246 ACTCATTCAGAAAACGCGAGATCAGCAG- 1299  
DB 1718 - 1733

QY 1300 GATGCGACCGAT- 1338  
DB 1734 GlyValThrAspProAspLeuTrpLeuAspIleArgAlaThrLeuArgSerLeuGluLys 1753

QY 1339 AAGTCTACGAGAACCAAGCGGTAAGGTTGAGAAGCTGCAAAATCCAAAGTGGC 1398  
DB 1754 ProValSer- 1766

QY 1399 AGAAG- 1449  
DB 1767 LysArgGlnValLeuHisLeuGluLysPheThrProAlaIleHisAla- 1782

QY 1450 AAATGTAATGGCCCGGCTGCTCACGTGGCGCAGCCGACTCGGTGTGACTGCTAGTAAT 1509  
DB 1782 - 1782

QY 1510 GACTGTATCTCTAAACACGCGCCAGCGCAATGAAGTTTCTAAGCTCAGTAAAGACAG 1569  
DB 1783 - 1785

QY 1570 AAGCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCAGAGAGCCAGTCTTCGGAATGC 1629  
DB 1786 HisSerGlnIleThrLysArgAlaArgLeuGluSerAspThrProSerLysAlaArgVal 1805

QY 1630 GGTGCTCAGCGCAGGTATTAAATCTCTGTGTGCAAGAGACCGAGTCCGAGAAAAA 1689  
DB 1806 SerLeuGlyArgSerAspSerIleSerThr- 1818

QY 1690 GAGACACAGTGAAGAGCGACTGTGTCTCGCGGAGTGAAGCACTCGG- 1743  
DB 1819 GluSerGlyAlaValProTyrAlaAlaProValProSerGluAlaValArgGlyIle 1838

QY 1744 - 1791  
DB 1839 ValProAlaLeuThrProAlaAlaAspSerProAlaSerArgSerGlyArgAsnAspAsp 1858

QY 1792 AATTCAATGCGTAAAGCCA- 1836  
DB 1859 SerPheAlaAlaSerProLeuTrpAspAlaLysThrGlyProSerPro 1876

RESULT 2  
US-09-252-991A-32447  
; Sequence 32447, Application US/09252991A  
; Patent No. 6531795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 32447  
; LENGTH: 700  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
; US-09-252-991A-32447

## Alignment Scores:

Pred. No.: 7,546-06 Length: 700  
 Score: 171.00 Matches: 150  
 Percent Similarity: 31.44% Conservative: 71  
 Best Local Similarity: 21.34% Mismatches: 276  
 Query Match: 3.65% Indels: 206  
 DB: 4 Gaps: 25

US-09-787-016A-1 (1-2610) x US-09-252-991A-32447 (1-700)

QY 188 GGTTCGTCGTCAGGGTTTTGGTGTATTAGGATTTCCAGGAAAAGTGTCCAGCTT 247  
 Db 19 GlyAspValCysGlnAlaLeuAlaGlnValGlyGln-GlnGlnAspArgAlaHi 38  
 QY 248 TCAGTGTGGAGCAGGTATGGACGCAAAAGCGCACCCGAGCAATAGGAGGACCATAGG 307  
 Db 38 sGlnCys-----AlaGlyGlnGlnArgHisProGlyArgAlaGly----- 51  
 QY 308 CCATCAACCCACCAGCAAGAGTTCAGGAAAACATGGGGTTTTCCAGAGGACCCTATCG 367  
 Db 51 ----- 51  
 QY 368 CCAGCGAGAGGG-----CGCAGGGGACGGGAGGCTGACCCACTGGAGCGCCACCCC 421  
 Db 52 -GlnAlaProGlyGlnLeuArgAlaGlyGlnGlyAspGluGlyAspGlyAlaGlyArgG1 71  
 QY 422 CACAGCAGCAGCTGGGCTCTCCCTCGCGCGGAGTGGGAGGCGCCCAAGCGCCTGAGC 481  
 Db 71 yGlyGlyAspGlyGlyHisAlaAspGlyAspGln----- 82  
 QY 482 GCGTGGAGCAGTCTCTGACCATTCGCGCGCGCGCGGAGGAGGAGCATGCTGTCTCCC 541  
 Db 83 -----AspHisLeuGlnAlaHisAlaLeuHisArgHisAlaGluLeuG1 97  
 QY 542 TGGAGGA-----TTCTGGTGAGCCCACTCTCTGCGCCCGCCAGCAG 580  
 Db 97 yGlyAlaGlyIleAlaGlnLeuGlnGlnPheGlnAlaThrValValGlnArgGlnAr 117  
 QY 581 ACGCCGAGACAGCTTCGAGGCGCGGCGTGAAGCGCTCTGAGACCAAGAGCGCCGCC 640  
 Db 117 Shis-----ValGlyGlnGlnAlaGlnArgCysArgAspAspValArgPro-- 132  
 QY 641 AGCTGCTTCCACAGCTGTGAAGGAACGACCCCTCTTCTGAAAAGGT-----GA 691  
 Db 133 -----GlyProGlyValG1 137  
 QY 692 AAGGAGGGATACACAGATGACACTC---CGATAGTCAGACGATGCGCTGACCTTGA 748  
 Db 137 nArgAlaGlyGlnProGluLeuGlyIleHisArgLeuValGluValAlaGlyAspGlnG1 157  
 QY 749 AAGAGCTTCAGAACTCGCTTCGACGAGCGGGACACAGGACCCACTGAGAG----- 800  
 Db 157 hGlnAlaValGlnAlaArgGlnAlaGlyAlaAspThrAspAlaHisGlnHisGlnAlaVa 177  
 QY 801 -----GCCCTGAAAGGGATCCAGATCGCTCGGGAGAGGAGAGCCCGGG 844  
 Db 177 lAlaAlaGlyGlyAlaProAlaArgGlnGlnHisArgGlnArgAlaGluAl 197  
 QY 845 AGGAGGCTCCGCGAGACTGTGGCTCGAGCGCCAGTCACACTGTGGAGGCGCTCTGCG 904  
 Db 197 aGlyGlyGlnArgArgGln-----ArgGlyG1 206  
 QY 905 CCAGTAAGCAGGAGCCCGAGAACGAT-----CAGGGGTTGTGCCAGGCTGGGA 955  
 Db 206 yGlnArgAlaGlyIle-GlnGlnAspAspLeuHisAspGlyAlaAlaGlyGlyAlaGlyG 226  
 QY 956 AAGATGACAGAGAGTAGTAAAGTGGAGGAAAGCGGCTCAGGACATCAAGATAGGAGC 1015  
 Db 226 lGlyGlnThrAspAspValGlyArgAlaGlnArgIleAlaArgAspArgLeuGluAspArgP 246  
 QY 1016 CTGGAGACTTGGCGGACCGAAGAGCTGAATGTGAGGGTTTACACCCCGCCCTGTATT 1075  
 Db 1075 ----- 1075

Db 246 roGlyHisAlaGlnArgGlyAlaAspGlnAlaGlyGlyGluHisProTrpGlnAlaTyra 266  
 QY 1076 GCATTTCGCCGCGACGCTCACAACAACAGGTTTATGATTCTGTGACCGCTGCGAAGAT 1135  
 Db 266 laLeu-----AspAspA 270  
 QY 1136 GGTTCATGCGGATGTGGGCAATTTCTGAGGCTCGAGGAGGCTTTTGTGAAAGGATG 1195  
 Db 270 laLeuGlyGlyGlyValLeuAlaLeuAlaGlnAlaGlyGluHisLeuCysArgAlaG 290  
 QY 1196 GGGAAAGACTATATCTGCCCAACTGCACCTTCGCAAGTCGAGATGAGACTCAITCAG 1255  
 Db 290 lnArgGlnAlaAlaGlyProGlnGlnArgArgGlnArgGlnSerAspProArgGluA 310  
 QY 1256 AAACGGCAGATCAGCAGAGCTAAATGAGACTCTGAGACTGCTGATGGCACCATTGTA 1315  
 Db 310 spGlnAlaGlyGlnArgArg---ProTrpProProThrGlnArgGlnGlyAlaGlnAlaG 329  
 QY 1316 CAAGTATAGGAACAATAGACGAGAAG----- 1341  
 Db 329 lnGlyValGluThrValHisArgArgLeuArgArgLysArgLeuThrArgThrMetLysA 349  
 QY 1341 ----- 1341  
 Db 349 laGlyProProSerSerLysAlaThrIleProAlaAlaIleSerAlaGlyAlaAlaSerS 369  
 QY 1342 -----TCTAGCGAAGACCAAGGGATA-----AAGGGTA 1369  
 Db 369 erArgProArgMetSerAlaProSerSerArgProAlaAlaIleSerAlaGluLysGlyA 389  
 QY 1370 GAATGTGAAGAGTCGCAATCCAAAGTGGCAAG-----AAGAAAC 1408  
 Db 389 rgIleGlnArgTrpSerGlyProAlaSerSerArgAlaArgCysGlyAlaThrArgProT 409  
 QY 1409 TCAAGATCTTCCACCT-----GTGATAGAGGCGCTGGTGCTCA----- 1449  
 Db 409 hrLysAlaIleGlyProAlaArgAlaValAlaAlaProAlaSerSerIleThrAlaArgG 429  
 QY 1450 -----AAATGTATTGCGCCCGGCTGTGTCTGACGTGGCGGACCCG 1489  
 Db 429 lnAlaSerSerGlnValArgArgSerThrProSerAlaArgProIleSerTrpProA 449  
 QY 1490 ACTCGGTGACTGCAATGACTGTCTCTCAACACGCGCGCAGCGCACANTGAAGTTTC 1549  
 Db 449 rgAlaArgAlaLeuSerLeuAlaAlaArgAlaLysAlaAlaAlaSerProAlaProMeta 469  
 QY 1550 TAAGCTCAGGT-----AAAGNACAGAGCCCAAGCCTAAAGAAAGAGATGAAGTGAAGC 1603  
 Db 469 laArgProGlyHisProSerArgAlaLysProLeuProGluSerGluProAlaAlaGlnT 489  
 QY 1604 CAGAGAAGCCAGTCTTCCGAAATGCGGTCT-----CAGCAGGTATTAAATCT 1654  
 Db 489 rpaGlyArgSerLysThrSerGlyGlyAlaPheSerThrMetIleProValSerGluA 509  
 QY 1655 CTTCGTGCAACAGAGACCGCTCCAGAAAAAAGAGACACACA-----GTGAAAGAGG 1708  
 Db 509 laSerIleThrAspLysProAlaProAlaSerThrArgArgThrGlyLeuValProValS 529  
 QY 1709 CAGTGGTGGTCCCTCGCGGAGTGAAGCAGCTCGGAAAGGAGAGCAGCTTGTGAGAGCACA 1769  
 Db 529 erArgAlaTrpProArgArgAsnThrAlaMetProAlaArgProAlaProArgLysAlaG 549  
 QY 1769 CGCGCTCG-----TGGCGGAGCGATCACAATTACATTACAGTAAGCCAGAAAGACTGTG 1825  
 Db 549 lnProArgTyTrpProMetLeuAlaArgSerAsnSerAlaThrProThrIleThrAlaAs 569  
 QY 1826 CTCCCTCGCGCTCACTGTGTATTAATATGATGATACCTAGGGGTTGGCCTCTCTGGACC 1885  
 Db 569 erAlaAlaProAlaLeuIle----- 576  
 QY 1886 CTTCCCGTCTTTCTGGATAGCCATCCCTGGGCTGTCCAGGACTGGAGGTTCGAGCTT 1945  
 Db 576 roSerArg-----ProGlySerAlaSerGlyLeuArgValSerAla- 589



QY 1946 TGTGTAAAGTGCATCACAGACACCGGCTGCACCATCACCGGGAAGCAGAGCCCATGTCCA 2005  
 Db 590 Cysile-SerAlaProAlaThrProSerAlaAlaProAlaSerSerAlaThrArgValar 609  
 QY 2006 G 2006  
 Db 609 g 609

## RESULT 3

US-08-801-308-1  
 ; Sequence 1, Application US/08801308  
 ; Patent No. 6368790  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Scott's, Robert E.  
 ; TITLE OF INVENTION: CDNA ENCODING P2P PROTEINS AND USE OF  
 ; TITLE OF INVENTION: P2P CDNA-DERIVED ANTIBODIES AND ANTISENSE REAGENTS IN  
 ; DETERMINING THE PROLIFERATIVE POTENTIAL OF NORMAL,  
 ; TITLE OF INVENTION: ABNORMAL AND CANCER CELLS IN ANIMALS AND HUMANS  
 ; NUMBER OF SEQUENCES: 4  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESS: Weiser & Associates, P.C.  
 ; STREET: 230 S. Fifteenth Street, Suite 500  
 ; CITY: Philadelphia  
 ; STATE: PA  
 ; COUNTRY: USA  
 ; ZIP: 19102  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC Compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/801,308  
 ; FILING DATE: 18-FEB-1997  
 ; CLASSIFICATION: 514  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Weiser, Gerard J.  
 ; REGISTRATION NUMBER: 19,763  
 ; REFERENCES/DOCKET NUMBER: 372,6435P  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 215-875-8383  
 ; TELEFAX: 215-875-8394  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1404 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-801-308-1

Alignment Scores:  
 Pred. No.: 1,53e-05 Length: 1404  
 Score: 169.00 Matches: 140  
 Percent Similarity: 32.68% Conservative: 77  
 Best Local Similarity: 21.08% Mismatches: 238  
 Query Match: 3.60% Indels: 210  
 DB: 4 Gaps: 29

US-09-787-016A-1 (1-2610) x US-08-801-308-1 (1-1404)

QY 262 GGTATGAGCAGCAAGGCGACCGCATGAGGCGACCTAAGCCCATCAAAACCCACC 321  
 Db 829 GlyProGlnGluLysValAspGlyAspArgGluLysSerPro-Arg---SerGluProPr 847  
 QY 322 AGCAAGAGTTTCAGAAACATCGGGTTTCGAAAGGACCACTATCCCAAGCGAGAGGCG 381  
 Db 847 oLeuLysLysAlaLysGluLysGluLysAlaThrLysIleAspSerValLysProSerSerSe 867  
 QY 382 GAGGGGACCGGAGGTCACCCACTGGAGCGCCACCCACAGCAGGAGCTGGGCGCTG 441  
 Db 867 rGlnLysAspGluLysValThrGlyThrProArgLysAlaHisSerLysSerAlaLysAs 887

QY 442 -----TCCCTGGCGCGCAGTGGGAGG----- 462  
 Db 887 pThrArgGlnSerGlnProArgThrArgArgSerLysArgThrValProLysThrSe 907  
 QY 463 -----CAGCCCAAGCGCAGTGCAGCGGTGAGCAGTTCCTGACCAATT----- 504  
 Db 907 rSerGlnLysSerGlnProValArgThrArgProArgSerIleuArgLysIleAsnTy 927  
 QY 505 -----CGCGGCGCGCGCAGCAGGAGCATGCTGTCTCCCT 542  
 Db 927 rLeuIleAlaArgGluLysAsnGluArgGluLysArgLysLysSerValAspLysAspPh 947  
 QY 543 CGAGGATTCTGGT-----GAGCCCACTCTCTGCCCCCGCCACAGAGCGC 584  
 Db 947 eGluSerSerSerMetLysIleSerLysValGluGlyThrGluIleValLysProSerPr 967  
 QY 585 CGACAGCCTCCGAGGCGCCTGGAAGCGCTTCTGAGCAGCAGAGCGCCCCCAGTC 644  
 Db 967 oLysArgLysMetGluGlyAspValGluLysLeuGluArgThr-----ProGluLys 984  
 QY 645 TGCTTCCACAGCTGTGAAGGAGCAGCAGCGCTCTTCTGAAAGAGTGAAGAGGGGATGA 704  
 Db 984 sAspLysIleAlaSerSerThrThrProAla-----LysLysIleLysLeuAsnArgL 1002  
 QY 705 CCACGATGACACTCCGATAGTGACCGGATGGCTGACCTTGAAGAGCTTCAGATCG 764  
 Db 1002 u-----ThrGlyLysLysIleGlyAsnAl 1010  
 QY 765 CTTTCGCGAAGCGGGAACAGAGCGCCACTGAGAGCGCCCTGAAAGGATCCAGAGTCG 824  
 Db 1010 aGluAsnAlaSerThrThrLysGluProSerGluLys---LeuGluSerThrSerSerLy 1029  
 QY 825 CCTG-----CGAAGAGCGCGGAGGAGGGTCC 854  
 Db 1029 sIleLysGlnGluLysValLysGlyLysAlaLysArgLysValAlaGlySerGluLys 1049  
 QY 855 CGCGAGACTCTG----- 867  
 Db 1049 rSerSerThrLeuValAspTyrThrSerThrSerThrGlySerProValArgLy 1069  
 QY 868 -GGCTCCGAGCGCCAGTGCAGTGGAGGGGCTCTGCGCCAGTAAAGAGGAGCGCGAA 926  
 Db 1069 sSerGluGluLysThrAspThrLysArgThrValIleLysThrMetGluGluTyAsnAs 1089  
 QY 927 CGAT-----CAGCGGGTGTGTCCAGCGC 950  
 Db 1089 rAspAsnThrAlaProAlaGluAspValIleMetIleGlnValProGlnSerLysTr 1109  
 QY 951 TGGGAAGATCACAGAGAGTAAAGTTGGAGGAAAGCGCGCTCAGGACATCAAGATGA 1010  
 Db 1109 pAspLysAspPheGluSerGluGlu-----AspValLysThrTh 1124  
 QY 1011 GGAGCCT---GGAGACTTGGCGCCAGCG-----AAGCC 1040  
 Db 1124 rGlnProIleGlnSerValGlyLysProSerSerIleIleLysAsnValThrLysPr 1144  
 QY 1041 T-----GAATGAGGGTTTACGACCCCAACCGCCTGTATTCGAT 1079  
 Db 1144 cSerAlaThrAlaLysTyrThrGluLysGluSerGluGlnProGluLysLeuGlnLysLe 1164  
 QY 1080 TTGCGCGCAGCTCACAAACACAGTTTATGCTGTGTCACCGCTGTGAGAGATGTT 1139  
 Db 1164 uProLysGluLysSerHis-----GluLeuMetGlnHisGluLe 1177  
 QY 1140 TCATGGCGATTGTGGCGCATTTCTGAGGCTCGAGGAGGCTTTTGAAGG-----AA 1193  
 Db 1177 uArgSerSerLysGlySerAlaSerSerGluLysGlyArgAlaLysAspArgGluHis 1197  
 QY 1194 TGGGGAAGCATATATCTCCCAAC-----TG 1220  
 Db 1197 rGlySerGluLysAspAsnProAspLysArgLysSerGlyAlaGlnProAspLysGluSe 1217  
 QY 1221 CACCATTTCTGCAGATGCAGGATGAGACTCATTCAGAAACGGCAGATCAGCAGGAGCTAA 1280

Db 1217 rThValAspArgLeuSerGluGlnGlyHisPhelYsThrLeuSerGlnSerSerLysG1 1237  
QY 1281 ATGAGACCTGGAGATGCTGATGGCCAGGATGTACAGTATAGGAAACAATAGAGCAGAA 1340  
Db 1237 uThrArg-----Th 1240  
QY 1341 GTCTACGAGAACCAAGGATTAAGGTTAGATTGAGAAAGCTGCAATCCAGTGGCAA 1400  
Db 1240 rSerGluLysHisGluSerValArgGlySerSerAsnLysAspPheThrProGlyArgAs 1260  
QY 1401 GAAGAACTCAAGATCTTCCAGCCTGTGATAGAGCGCTGGTGCCTCAAAATGATTGG 1460  
Db 1260 pLysLysValAsp----- 1264  
QY 1461 CCCCGGTGCTGACGTGGCGCAGCCGACTCGGTGTACTGCAATGATGATGATCT 1520  
Db 1265 -----TyrAspSerArgAspTyrSerSe 1272  
QY 1521 CAACACCGCGCAGCAGCAATGAAGTTTCTAAGCTCAGGTAAAGACAGACAGCAAGCC 1580  
Db 1272 rSerLysA-gAAspGluArgGlyGluLeuAlaArgArgLysAspSerProProArgG1 1292  
QY 1581 TAAAGAA-----AAGATGAAGATGAAGCCAGAGAACCCAGTCTTCCGAA 1625  
Db 1292 yLysGluSerLeuSerGlyGlnLysSerLysLysLeuArgGluArg---AspLeuProLy 1311  
QY 1626 ATCCGCTGCTCAGCAGGTATTAATCTCTTCTGTG-----CACAAGAGCCAGCTCC 1679  
Db 1311 sLysGlyAlaGluSerLysLysSerAsnSerProProArgAspLysLysProHisAs 1331  
QY 1680 AGAAAAAAGACACACAGTGAAGAGGAGGAGTGTGTCTCTCTG-----CAGCAGGAGTGAACACT 1739  
Db 1331 pHisLysAlaProTyrGluThrLysArg----- 1340  
QY 1740 CGGGAAGGAGCAGCTGTGTGAGCAGCAGCGCGTGTGTGGGAGAGGATCACAATTACAA 1799  
Db 1341 -----ProCysGluGluThrLysPro-----ValAspLysAsn---Se 1352  
QY 1800 TGCAGTAAAGCCCAAGAAAGACTGCTGCTCCTCG-----CCGTCACCTGTGTA 1847  
Db 1352 rGlyLysGluArgGluLysHisAlaAlaGluAlaArgAsnGlyLysGluSerSerGlyAl 1372  
QY 1848 TAAATGTATATACACTAGGCTGTGGCTCTCGACCCCTCCGCTCTTCTTGTGATAGC 1907  
Db 1372 aAsnCysHisValTyrLeuThrArgGlnThrLeu---ProTyrArgArgSerTyrLeuLe 1391  
QY 1908 CATCCCTCG 1917  
Db 1391 uGlyArgTyr 1394

RESULT 4  
US-09-252-991A-18387  
; Sequence 18387, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 18387  
; LENGTH: 1075  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-18387

Alignment Scores: 3.24e-05 Length: 1075  
Score: 164.50 Matches: 126  
Percent Similarity: 34.53% Conservative: 76  
Best Local Similarity: 21.54% Mismatches: 228  
Query Match: 3.51% Indels: 155  
DB: 4 Gaps: 26

US-09-787-016A-1 (1-2610) x US-09-252-991A-18387 (1-1075)

QY 283 CCGACCAATGAGGAGGACCTTAAGCCATCAACCCACAGCAAGAGTTTCAGGAAACA 342  
Db 553 ProGlnHisThrGluAlaProValGluProAlaLysProMetProGluProSerLeuPhe 572  
QY 343 TGGGCTTTTGAAGACCACTATATCCCAAGCGAGAGGCGCAGGAGCGCGAGGCTGAC 402  
Db 573 GlnGlyLeuValLysSerLeuValGlyLeuPheAlaGlyLysAspGlnProAlaLys 592  
QY 403 CCACCTGGAG-----CCGCCACCCCCACAGCAGAGCTGGGCTCTCCTCGCGGCGAGT 456  
Db 593 ProAlaGluThrSerLysProAlaAlaGluArgGlnThrArgGlnAspGluArgAsn 612  
QY 457 GGGAGGAGCCCGACGAGCGGCTGAGCGGCTGAGCAGTTCCTGACCATTCGCGGCGCGCC 516  
Db 613 GlyArgGlnGlnAsnArgArgAspGlyArgAsp-----GlyAsnArgArg 628  
QY 517 GCGAGGAGGAGCATGCTGTCTCTCCCTGAGGATTCGTGTAGCCACCGCTCTCTCCCGCCG 576  
Db 628 AspGluGluArgLysProArgGluAlaArgGlnProArgGlu----- 645  
QY 577 ACAGACCGCGAGACACTTCGAGGCGGCTGGAAGCGCTTCTGAGACGAGAGCGCG 636  
Db 646 -----GluArgAlaGluArgProAsnArgGluArgSerGluArgArgGlu 662  
QY 637 CCCAGTCTGCTCCACAGCTGTGAAGGAACGACAGCCTCTCTGAAAGAGTGAAAGGA 696  
Db 663 GluArgAla-----GluArgProAlaArgGluGluArg----- 673  
QY 697 GGGGATGACACGATGACACCTCCGATAGTGAACGATGGCTCTCACTTGAAGAGCTT 756  
Db 673 ----- 673  
QY 757 CAGATCCGCTTCGAGGAGCGGGAACAGGAGCCCTGAGAG-----CCC 804  
Db 674 GlnProArgGluGlyArgGluGluArgAlaGluArgThrProArgGluGluArgGlnPro 693  
QY 805 CTGAAGGGATCCAGAGTCCG-----CTGGGGAAGAGCGCGGAGAGGCTCCCGCGAG 861  
Db 694 ArgGluGlyArgGluGlyArgGluGluArgSerGluArgArgGluGluArgAlaGlu 713  
QY 862 ACTGTGGCTCCGAGCGCAGTGCACACTGTGAGGCGCTCTCCCGAGTAGCAGGAGCC 921  
Db 714 ArgProAlaArgGluGluArgGlnProArgGluGlyArgGluGluArgValGluArgPro 733  
QY 922 GAGAAGCATCAGGGGTTGTGTCCAGCTGGGAAGATGACAGAGAGAGTAAG----- 975  
Db 734 AlaArgGluGlu-----ArgGlnProArgGluAspArgGlnAlaArgAspAla 749  
QY 976 -----TTGGAGGGAAGGCG-----GCTCAGACATCAAGAT 1008  
Db 750 AlaAlaLeuGluAlaGluAlaLeuProAsnAspGluSerLeuGluGlnAspGluGlnAsp 769  
QY 1009 GAGGAGCTGAGACTTGGCGCGCAGCGCTGAACTGTAATGTAGGCTTACGACCCCAAGCC 1068  
Db 770 AspThrAspGlyGlu-----ArgProArgArgArgSerArgGlyGln----- 783  
QY 1069 CTGTATTGCAATTGCGCCAGCTCACAACACAGGTTTATGATTTGTGTGACCGCTGT 1128  
Db 784 -----ArgArgArgSerAsnArgArg-----GluArgGln 793  
QY 1129 GAAGAAATGTTTCATGGCGATGTGTGGCATTTCTGAGGCTTCGAGGAGGCTTTTGGAA 1188  
Db 794 ArgGlu---ValSerGlyGluValGluGly---SerGluAlaThr----- 806

```

QY 1189 AGGAATGGGGAAGACTATATCTGCCCCAACTGCACCATTT----- 1227
Db 807 -----AspAlaAlaAlaProLeuAsnThrValAlaAlaAlaAlaAlaGly 823
QY 1228 CTGCAAGTCAGGATGAG-----ACTCATTC 1254
Db 824 IleAlaValAsnSerGluAlaValGluAlaAsnValGluGlnAlaProAlaThrSer 843
QY 1255 GAAACGGCAGATAGCAGGAAGCTAAATCGAGACTGGAGATGCTGATCGCACCATTTGT 1314
Db 844 GluAlaAlaSerGluThrThrAla-----SerAspGluThrAspAla 857
QY 1315 ACAAGTATAGGACATAGACAGAGTCTAGCGAGACCAAGGGATTAAGGGTAGA--- 1371
Db 858 SerThrSerGluAlaValGluThrGlnAspAlaAspSerGluAlaAsnThrGlyGluThr 877
QY 1372 -----ATTGAGAAGCTGCAAAATCCAAAGTGGCGCAAGAGAAA 1407
Db 878 AlaAspIleGluAlaProValThrValSerValValArgAspGluAlaAspGlnSerThr 897
QY 1408 CTCAGATCTTCAGCTGTGTAGAGCGCTGGTGGCTCAAAATGTATT----- 1458
Db 898 LeuLeuValAlaGlnAlaThrGluGluAlaProPheAlaSerGluSerValGluSerArg 917
QY 1459 -----GGCCCC 1464
Db 918 GluAspAlaGluSerAlaValGlnProAlaThrGluAlaAlaGluGluValAlaAlaPro 937
QY 1465 GGGTGTGTGTCAGTGGCGAGCCGACTGGTGTACTGCGAGTAATGACTGTATCTCCANA 1524
Db 938 ValProValGluValAlaAlaProSerGluProAlaAlaThrGluGluProThrProAla 957
QY 1525 CAGCCGCGCA-----GGACAATGAAGTTTCTAAGCTCAAGTAAAGAACAGAG 1572
Db 958 IleAlaAlaValProAlaAsnAlaThrGlyArgAlaLeuAsnAspProArgGluLysArg 977
QY 1573 CCAAAGCCTAAAGAAAGATGAAGTAAAGCCAGAGAGCCCGCTTCGGAAATCGGT 1632
Db 978 ArgLeuGlnArgGluAlaGluArgLeuAlaAlaArgGluAlaAlaAlaGluAlaAla 997
QY 1633 GCTCAGCGAGGTATTAATCTCTCTGTGCACAGAGACAGCTCCAGAAAAAAGAG 1692
Db 998 AlaGlnAlaAla-----ProAlaValGluGluIleProAlaValAlaSerGluGlu 1014
QY 1693 ACCACAGTGAAGAGGAGTGGTGGTCCCTGCGCGAGTGAAGCACTCGGAAAGGAAGCA 1752
Db 1015 AlaSerAlaGlnGluGlu-----ProAlaAlaProGlnAla-----GluGlu 1028
QY 1753 GCTGTGAGAGCAGCAGCGCTGTGGCGAGCGGATCACATACATGAGTAAAGCCA 1812
Db 1029 IleThrGlnAlaAspValProSerGlnAla---AspGluAlaGlnGluAlaValGlnAla 1047
QY 1813 GAAAAGACTGTGCT 1827
Db 1048 GluProGluAlaSer 1052

```

## RESULT 5

```

US-09-252-991A-27005
; Sequence 27005, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE OF INVENTION: ABRUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: 1999-02-18
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142

```

```

; SEQ ID NO 27005
; LENGTH: 1427
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27005
Alignment Scores:
Pred. No.: 3,7e-05 Length: 1427
Score: 164.50 Matches: 111
Percent Similarity: 36.18% Conservative: 50
Best local Similarity: 24.94% Mismatches: 128
Query Match: 3.51% Indels: 156
DB: 4 Gaps: 25

```

US-09-787-016A-1 (1-2610) x US-09-252-991A-27005 (1-1427)

```

QY 273 CAAGGCGAGCCGACGCAATGAGGAGGACCTTAAGCCATCAACCCACCACGAAAGAGTT 332
Db 770 ArgArgGlnProGluPro---GlyAlaValProGlyArgAlaAlaHisArg----- 785
QY 333 CAGGAAACATGGGGTTTTTCGAGGACCATCTATCCCAAGCAGAGGGCGCAGGGGAGCG 392
Db 786 -----GlyAspProTyrLeuArgProArgHisArgArgGlyArg 798
QY 393 GGAGGCTGACCATCTGGAGCGCCACCCACAGCAGCAGCTGGGCTGTCTCCGCGCG 452
Db 799 AlaArgGlyHisAlaGlyAlaAla---ProHisGlnProAlaGly-----AlaAla 814
QY 453 CAGTGGGAGGAGCGCCAGCGCACTGAGCGCGTGGAGCAGTTCTCTACANTTGGCGCG 512
Db 815 LeuLeuGlyAlaAlaGluAspHis---SerArgTyrArgAlaAlaAspHisProAlaGly 833
QY 513 CCG-----CGCAGGAGGAGCGCTGCC----- 533
Db 834 AlaAspArgLeuSerLeuArgArgArgArgHisAlaHisAlaArgThrAlaLeuHis 853
QY 534 -----TGCTCTCCTT-----GGAGGATTC 551
Db 854 ProAlaAlaValArgArgLeuProAlaLeuAlaArgGluGlnArgThrLeuGlyGly--- 872
QY 552 TSGTGAGCCACGCTCTCCCGCCACAGAGCCGAGCAGCAGCTCCGAGGGCAGCGTGA 611
Db 873 -----AlaGluLeuValProHisHisArgArgGlyArgHisArgProThrLeu 890
QY 612 AAGCGCTTCTGAGACCAGAGCGGCCCGCTCTCCACAGCTGTGAAGAGAGCA--- 669
Db 891 LysArg-----ArgArgPro-SerArgSerValSerGlnArgArgAl 905
QY 670 -----CCAGCCTCTTCTGAAGAGTGAAGAGGGAGTACCCAGATGACACCTC 719
Db 905 aLeuAlaThrProArgThrGluGlnArgGlnHisGlyGlyArgAspAspGluHisArg 925
QY 720 CGATAGTGACGCGATGCGCTACCTTGAAGAGCTTCAGAACTCGCTTCGAGGAAGCG 779
Db 925 gGluArgProIleGluGly-----GlnArgLeuProAlaValVa 938
QY 780 GGAACAGAGCGCCACTGAGAGGCGCCCTGAAGAGGATCCAGAGTCGCTCGGAAAGAGCG 839
Db 938 lGluGlnGluProGlyHis-----GlnArg 946
QY 840 CCGGAGGAGGGGTTCGCGCGAGACTGTGGCTCCGAGGCCAGTGAC-----ACTGT 890
Db 946 gArgGluAspArgProAlaAlaThrAspAlaAspProProAlaAspProGlyGlyThrHi 966
QY 891 GGAGGCGCTCTGCGCCAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCAGCG 950
Db 966 sArgSerValValGluAlaArgArg-----GlnGlyIle-----GlnAl 979
QY 951 TGGG-----AAGATCAGCAGAGAGTAAAGTTGGAGGAAAGCGG--- 990
Db 979 aGlyHisAlaGlyValGlyAlaLysProAspGlnSerGlnArgGlnGlyArgAlaGl 999
QY 991 -----GCTCAGGACATCAAGATGAGGAGCGCTGGAGAC----- 1023

```

```

Db      999 yclnProGlyAlaGlnArgGluGlnSerGlyGlyLysGlnHisGluGlnAr 1019
QY      1024 -----TTGGCCCGCAGCAGGAGCCTGAAATGTGAGGTTACGA 1058
Db      1019 gGlnHisProPheGlnAlaLysAlaArgGlyLysProAlaGluGlnGlyThrGluAs 1039
QY      1059 CCCCAGGCCCTGTATTGCATTTGCCGCCAGCCTCACACACAGGTTTATGATTGCTG 1118
Db      1039 pPro-----1040
QY      1119 TGACCGCTGTGAAGAATGGTTTCATGCGCATTTGTGGGCATTTCTGAGCTCGAGGG-- 1176
Db      1041 -AlaAspAlaGlnGlnGlyAlaLysAlaArgGlyLysMetGlyGlyGlnAlaGlyLys 1060
QY      1177 -AGCTTTTGGAAAGAAATGGGAGAGACTATATCTGCCCAACTGCACCATTTCTGCAAGT 1235
Db      1060 pGlnGlnLeuArgArgProAlaHisAsp-----GluVa 1071
QY      1236 GCAGGATGAGACTCATTTACAGAAACGCCAGATCAGCAGGAGGCTAAATGAGACCTGGAGA 1295
Db      1071 lGluAlaGluHisHisGlnGlu-----LysGlnArgProGluGln 1084
QY      1296 TGCTGATGCGACCGGATTGTACAGTAGTAGGAAACAATAGACAG--1338
Db      1084 nGlnArgGly-----ThrSerLeuGlyAsnAlaLysGlnProGlyGluAlaAlaTh 1101
QY      1339 -----AGCTTAGCAGACCAAGGATTAAGGTAGATTGAGAAAGCTGCAATCC 1391
Db      1101 rValLeuArgSerAlaThrGlyArgAlaArgGlnPheArgGluArgAlaAlaGlnAl 1121
QY      1392 AAGTGGCAAGAG 1404
Db      1121 aProGlyGlnArg 1125

RESULT 6
US-08-769-309A-5
Sequence 5, Application US/08769309A
Patent No. 5741890
GENERAL INFORMATION:
APPLICANT: Scott, John D.,
APPLICANT: Nauert, Brian J.,
APPLICANT: Klauk, Theresa M.
TITLE OF INVENTION: Protein Binding Domains of Gravin
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower/233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/769,309A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: No. 5741890and, Greta B.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 27866/33451
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1780 amino acids
TYPE: amino acid

```

```

; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-769-309A-5
Alignment Scores:
Pred. No.: 4.97e-05 Length: 1780
Score: 163.50 Matches: 136
Percent Similarity: 30.71% Conservative: 79
Best Local Similarity: 19.43% Mismatches: 224
Query Match: 3.49% Indels: 261
DB: 1 Gaps: 29
US-09-787-016A-1 (1-2610) x US-08-769-309A-5 (1-1780)
QY      226 TCAGGGAAGTGTCCAAAGCTTTCAGTCTTGGAGCAGGTATGAGCACAAGAGCGACCG 285
Db      257 SerGlyGlnAlaValGluGlu-----CysLysGluGluGlyGluGlu 270
QY      286 AGCAATGAGGAGGCACCT-----AAGCCATCAAAACCCACGAG--324
Db      271 LysGlnGluLysGluProSerLysSerAlaGluSerProThrSerProValThrSerGlu 290
QY      325 -----AAGAGTTTCAGAAACATCG---GGTTTTCAGAGGACCAT 363
Db      291 ThrGlySerThrPheLysLysPhePheThrGlnGlyTrpAlaGlyTrpArgLysLysThr 310
QY      363 -----363
Db      311 SerPheArgLysProLysGluAspGluValGluAlaSerGluLysLysGluGlnGlu 330
QY      364 -----ATCGCCAAAGGAGGAGCGGCA 384
Db      331 ProGluLysValAspThrGluGluAspGlyLysAlaGluValAlaSerGluLysLeuThr 350
QY      385 GGGGACGCGGAGGCTGACCCACTGAGCGCGCA-----417
Db      351 AlaSerGluGlnAlaHisProGlnGluProAlaGluSerAlaHisGluProArgLysSer 370
QY      418 -----CCCCACAGCAGCAGCTGGGCTGTCCCTGGCGGCGC 453
Db      371 AlaGluTyrgLysValGluLeuProSerGluGluGlnVal-----384
QY      454 AGTGGAGGCGAG-----CCCAAGCGCAGCTGAGCGGTGGAG 489
Db      385 SerGlySerGlnGlyProSerGluGluLysProAlaProLeuAlaThrGluValPheAsp 404
QY      490 CAGTTCTCAGCATTGCGCGCGCGCGCGCAGGAGCATGCTGTCTCTCCCTGGAGGAT 549
Db      405 GluLysIleGluValHisGlnGluValValAlaGluValHisValSerThrValGlu 424
QY      550 TCTGTGAGCCCGCTCTCTGCCCCCGCCACAGACGCG-----GAGACAGCCTCCGAGGCGAGC 606
Db      425 -----GluArgThrGluGluGlnLysThrGluValGluGluThrAlaGlySerValPro 442
QY      607 GTGGAAGCGCTTCTGAGACCAAGCGCGCCAGTCTGCTTCCACAGCTGTGAAGGAA 666
Db      443 AlaGluGluLeuValGlyMetAspAlaGluProGlnGluAlaGluProAla---LysGlu 461
QY      667 CGACAGCCTCTTCTGAAAGGTGAAGAGGGGATGACACGATGATGATCCGATAGT 726
Db      462 LeuValLysLeuLysGluThrCysValSerGlyGluAspProThrGlnGlyAlaAspLeu 481
QY      727 GACAGCGATGGCTGACCTTGAAGAG-----753
Db      482 SerProAspGluLysValLeuSerLysProProGluGlyValValSerGluValGluMet 501
QY      754 CTTTCAAGATCGCTTTCGAGGAAGCGGGAACAGAGGAGCC-----ACTGAG 798
Db      502 LeuSerSerGlnGluArgMetLysValGlnGlySerProLeuLysLysLeuPheThrSer 521
QY      799 AGGCCCTTGAAGGGATCCAGATGCTGCTCGGAAGAGAGCGCGGAGGAGGCT-----852
Db      522 ThrGlyLeuLysLysLeuSerGlyLysGlnLysGlyLysArgGlyGlyAspGlu 541

```

853 QY -----CCCGCCGAGACTGTGGCTCCGAGGCCAGTGCACACT 889  
542 Db GUserGlyGluHisThrGlnValProAlaAspSerProAspSerGlnGluGlnLys 561  
889 QY GTGGAGGGCGTCTGCGCCAGTAAGCAGGAGCCGAG----- 924  
562 Db GlyGluSerAlaSerSerProGluGluProGluGluLeuThrCysLeuGluLysGly 581  
925 QY -----AACGATCAGGGGGTGTGCTCCAGGCTGGGAAAGATGACAGAGACT 972  
582 Db LeuAlaGluValGlnGlnAspGlyGluAlaGluGlyAlaThrSerAspGlyGluLys 601  
973 QY AAGTTGGAGGA----- 984  
602 Db LysArgGluGlyValThrProTrpAlaSerPheLysLysMetValThrProLysLysArg 621  
984 QY ----- 984  
622 Db ValArgArgProSerGluSerAspLysGluAspGluLeuAspLysValLysSerAlaThr 641  
985 QY -----AAGCGGCTCAGGACATCAAGATGAGGAGCTCGAGACTTGGGC 1029  
642 Db LeuSerSerThrGluSerThrAlaSerGluMetGlnGluMetLysGlySerValGlu 661  
1030 QY CGACCGAAGCCTGATGTGAGGTTACGACCCACGCCCTGTATTGTCATTTGCCGCCAG 1089  
562 Db GluProLysProGlu-----Glu 667  
1090 QY CTTCAACACACAGGTTTATGATTCTGTGACCGCTGTGAAGATGGTTTCATGCGAT 1149  
668 Db ProLysArgLys-----ValAspThrSerValSerTrpGluAlaLeuIle 682  
1150 QY TGTGTGGGATTTCTGAGGCTCGAGGAGCTTTGGAAAGGAATGGGAAGCATATATC 1209  
683 Db CysValGlySerSerLysLysArgAlaArg-----ArgArgSerSerAsp----- 698  
1210 QY TGCCCAAACTGCACATTTCTGAAGTGCAGATGAGACTCAITTCAGAAACCGCGATCAG 1269  
699 Db -----GluGluGlyGlyProLysAlaMetGlyGlyAspHis 710  
1270 QY CAGAAAGCTAAATGAGACCTGGAGATGCTGAT---GGCACCATTTGACAGTATAGGA 1326  
711 Db GlnLysAlaAspGluAlaGlyLysAspLysGluThrGlyThrAspGlyLeuAlaGly 730  
1327 QY AACTAGACAGAGTCTACCGAAGACCAAGGATAGAGGTAGAAATTCAGAAAGCTGCA 1386  
731 Db Ser-----GlnGluHisAspProGlyGlnGly---SerSerSerProGluGlnAlaGly 747  
1387 QY ATCCCAAGTGGCAAGAAACTCAGATCTTCAGCCTGTGATAGGCGCCTGCTGCC 1446  
748 Db SerProThr-----GluGlyGluGlyVal 755  
1447 QY TCAAAATGATATGCCCCGGGTGTGTACGTGGCGCGCAGCTCGGTACTCAGT 1506  
756 Db Ser----- 756  
1507 QY ATGACTGTATCTCAAAACGCGCGCGACATGAGTTCTAAGCTCAGGTAAAGAA 1566  
757 Db -----ThrTrpGluSerPheLysArgLeuValThr----- 766  
1567 QY CAGAAGCAAAAGCCTAAAGAAAGATCAAGATGAGCCAGAGAGCCAGCTCTTCGAAA 1626  
767 Db -----ProArgLysLysSerLysSerLysLeuGluGluLysSerGluAspSer 782  
1627 QY TGGGTGCTCAGGAGGTATTAATCTCTTGTGTGCACAAAGACAGCAGCTCCAGAAAA 1686  
783 Db Ile-----AlaGlySerGlyValGluHisSerThrProAspThrGluProGlyLysGluGlu 801  
1687 QY AAGAGACCACTGAAAGAGGCGAGTGTGTCTCCGCGGAGTGAAGCACTCGGGAAG 1746  
802 Db SerTrpValSerIleLysPheIleProGlyArgArgLysLysArgProAspGlyLys 821

1747 QY -----GAAGCAGCTTGTGAGACGACGCGCTGTGGCGGAGCCATCAAT----- 1794  
822 Db GlnGluGlnAlaProValGluAspAlaGlyProThrGlyAlaAsnGluAspAspSerAsp 841  
1795 QY -----TACATCAGCTAAAGCCAGCAAAAGACTGTCTGCT 1827  
842 Db ValProAlaValValProLeuSerGluTyAspAlaValGluArgGluLysMetGluAla 861  
RESULT 7  
US-08-994-570-5  
; Sequence 5, Application US/08994570  
; Patent No. 6090929  
; GENERAL INFORMATION:  
; APPLICANT: Scott, John D.,  
; APPLICANT: Nauert, Brian J.,  
; APPLICANT: Klauk, Theresa M.  
; TITLE OF INVENTION: Protein Binding Domains of Gravin  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower/233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/994,570  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: NO. 6090929and, Greta E.  
; REGISTRATION NUMBER: 35,302  
; REFERENCE/DOCKET NUMBER: 27866/33451  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1780 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-994-570-5  
Alignment Scores:  
Pred. No.: 4,978-05 Length: 1780  
Score: 163.50 Matches: 136  
Percent Similarity: 30.71% Conservative: 79  
Best Local Similarity: 19.43% Mismatches: 224  
Query Match: 3.49% Indels: 261  
DB: 3 Gaps: 29  
US-09-787-016A-1 (1-2610) x US-08-994-570-5 (1-1780)  
QY 226 TCAGGAAAGTGTCAAGCTTTCAGTGTGGAGCAGGTATCGACGACAAAGGCGACCG 285  
Db 257 SerGlyGlnAlaValGluGlu-----CysLysGlnGluGlnGluGlu 270  
QY 286 AGCAATGAGGAGCGACCT-----AAGGCCATCAAAACCCACCAGC----- 324  
Db 271 LysGlnGluLysGluProSerLysSerAlaGluSerProThrSerProValThrSerGlu 290  
QY 325 -----AAAGAGTTCAGGAAAAATGG-----GGTTTCGAGGACCACT 363  
Db 291 ThrGlySerThrPheLysLysPheThrGlnGlyTrpAlaGlyTrpArgLysLysThr 310  
QY 363 ----- 363

```

Db 311 SerPheArgLysProLysGluAspGluValGluAlaSerGluLysLysLysGluGlnGlu 330
QY 364 -----ATGCCAAGCCAGAGCGCGCA 384
Db 331 ProGluLysValAspThrGluGluAspGlyLysAlaGluValAlaSerGluLysLeuThr 350
QY 385 GGGGACGGAGGCTGACCCACTGGAGCGCCA-----417
Db 351 AlaSerGluGlnAlaHisProGlnGluProAlaGluSerAlaHisGluProArgLeuSer 370
QY 418 -----CCCCACACAGCAGCTGGCGCTGCTCCCTGGCGCGC 453
Db 371 AlaGluThrGluLysValGluLeuProSerGluGluGlnVal-----384
QY 454 AGTGGAGGCGAG-----CCCAAGCGCACTGAGCGCGTGGAG 489
Db 385 SerGlySerGlnGlyProSerGluGluLysProAlaProLeuAlaThrGluValPheAsp 404
QY 490 CAGTTCCTGACCATTCGGCGCGCGCGGCGGAGGAGGATGCTGCTCCCTGGAGGAT 549
Db 405 GluLysIleGluValHisGlnGluGluValAlaGluValHisValSerThrValGlu 424
QY 550 TCTGGTGGCCCGCTGCTCCCGCGCACAGAGCC---GAGACAGCCTCCGAGGCGAGC 606
Db 425 -----GluArgThrGluGlnGlyThrGluValGluGluThrAlaGlySerValPro 442
QY 607 GTGGAAGCGCTTCTGAGACACAGAGCGCGCCAGTCTGCTCCAGCTGTGAAGAA 666
Db 443 AlaGluGluLeuValGlyMetAspAlaGluProGlnGluAlaGluProAla---LysGlu 461
QY 667 CGACCGCTCTTCTGAAAGGTGAAAGAGGAGGATGACCAACGATCACCTCCGATAGT 726
Db 462 LeuValLysLeuLysGluThrCysValSerGlyLysAspProThrGlnGlyAlaAspLeu 481
QY 727 GACAGCGATGGCTGACCTTGAAGAG-----753
Db 482 SerProAspGluLysValLeuSerLysProProGluGlyValValSerGluValGluMet 501
QY 754 CTTGAGATCGCTTCGACAGAGCGGACAGAGCGCC-----ACTGAG 798
Db 502 LeuSerSerGlnGluArgMetLysValGlnGlySerProLeuLysLysLeuPheThrSer 521
QY 799 AGGCCCCGTAAGAGGATCCAGATCGCTGCGGAAGAGCGCGGAGGAGGT-----852
Db 522 ThrGlyLeuLysLysLeuSerGlyLysLysGlyLysArgGlyGlyGlyAspGlu 541
QY 853 -----CCGCGGAGACTGTGGCTCCGAGGCGGAGTGCACACT 898
Db 542 GluSerGlyGluHisThrGlnValProAlaAspSerProAspSerGlnGluGlnLys 561
QY 889 GTGGAGGCGCTGCTGCCCGCTAAGCAGAGCGCGAG-----924
Db 562 GlyGluSerSerAlaSerProGluGluProGluGluIleThrCysLeuGluLysGly 581
QY 925 -----AACGATCAGGGGTGTGTCGCCAGCTGGGAAGATGACAGAGAGT 972
Db 582 LeuAlaGluValGlnGlnAspGlyGluAlaGluGlnGlyAlaThrSerAspGlyGluLys 601
QY 973 AAGTTGGAGGCA-----984
Db 602 LysArgGluGlyValThrProThrAlaSerPheLysLysMetValThrProLysLysArg 621
QY 984 -----984
Db 622 ValArgProSerGluSerAspLysGluAspGluLeuAspLysValLysSerAlaThr 641
QY 985 -----AAGGCGCTCAGACATCAAGATCAGGAGCTGAGAGCTGGGC 1029
Db 642 LeuSerSerThrGluSerThrAlaSerGluMetGlnGluMetLysGlySerValGlu 661
QY 1030 CGACCGAGGCTGAATGTAGGGTTACAGCCCAACGCCCTGTATTGCTATTCGCCCGCAG 1089

```

```

Db 662 GluProLysProGlu-----Glu 667
QY 1090 CUTCACAAACACAGGTTTATGATTTCGTGTGACCCCTGTGAAGATGTTTTCATGGCGAT 1149
Db 668 ProLysAlaGlyS-----ValAspThrSerValSerTrpGluAlaLeuIle 682
QY 1150 TGTGTGGGCTATTCTGAGGCTCGAGGAGGCTTTTGGAAAGAAATGGGGAGACTATATC 1209
Db 683 CysValGlySerSerLysLysArgAlaArg-----ArgArgSerSerSerAsp 698
QY 1210 TGCCCAAACTGCACCATTCCTCAAGTCAGAGATGAGACTCATTTCAAGAACCGCAGATCAG 1269
Db 699 -----GluGluGlyGlyProLysAlaMetGlyLysAspHis 710
QY 1270 CAGGAAGCTAATGAGACCTGGAGATGCTGAT---GGCAGCATGTACAGATAGGA 1326
Db 711 GlnLysAlaAspGluAlaGlyLysAspLysGluThrGlyThrAspGlyLleLeuAlaGly 730
QY 1327 ACAATAGACAGAACTCTAGCGAAGCAAGGAGATAAAGGGTAGAATTCAGAAAGCTGCA 1386
Db 731 Ser-----GlnGluHisAspProGlyGlnGly---SerSerSerProGluGlnAlaGly 747
QY 1387 AATCAAGTGGCAGAGAACTCAAGATCTTCAGCCTGTGTATAGAGCGCTGTGGTCC 1446
Db 748 SerProThr-----GluGlyGluGlyVal 755
QY 1447 TCAAAATGTATTGGCCCCCGGTGCTGTACGTGGCGCAGCCGCTCGGTGTACTGCAGT 1506
Db 756 Ser-----756
QY 1507 AATGACTGTATCTCTCAACACCGCGAGCGACAATGAAGTTTCTTAAGCTCAGGTAAAGAA 1566
Db 757 -----ThrTrpGluSerPheLysArgLeuValThr 766
QY 1567 CAGAACCCAAAGCCTAAGAAAGATGAGATGAAGCAGAGAGCCAGTCTTCGGAAG 1626
Db 767 -----ProArgLysSerLysSerLysLeuGluGluLysSerGluAspSer 782
QY 1627 TCGCGTGTCTCAGCGAGTATTAAATCTCTCTGTGTCACAGAGACCACTCCAGAAAA 1686
Db 783 Ile---AlaGlySerGlyValGluHisSerThrProAspThrGluProGlyLysGluGlu 801
QY 1687 AAGAGACACACAGTGAAGAGGAGCAGTGGTGGTCCCTGCGCGGAGTGAAGCACTCGGGAAG 1746
Db 802 SerTrpValSerIleLysLysPheIleProGlyArgArgLysLysArgProAspGlyLys 821
QY 1747 -----GAAGCAGCTGTGTGAGAGCAGCAGCGCTGTGGCGGAGTCACAAT-----1794
Db 822 GlnGluGlnAlaProValGluAspAlaGlyProThrGlyAlaAsnGluAspSerAsp 841
QY 1795 -----TACAATGCAAGTAAAGCCAGAAAGACTGTGCT 1827
Db 842 ValProAlaValProLeuSerGluThrAspAlaValGluArgGluLysMetGluAla 861

```

## RESULT 8

US-07-853-913-2

Sequence 2, Application US/07853913

Patent No. 5338839

GENERAL INFORMATION:

APPLICANT: McKay, Ronald D.G.

APPLICANT: Lendahl, Urban

TITLE OF INVENTION: Nestin Expression As An Indicator of

TITLE OF INVENTION: Neuroepithelial Tumors

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hamilton, Brook, Smith &amp; Reynolds, P.C.

STREET: Two Militia Drive

CITY: Lexington

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02173

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/853,913
FILING DATE: 19920319
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/660,412
FILING DATE: 22-FEB-1991
APPLICATION DATA: US 07/603,803
FILING DATE: 25-OCT-1990
APPLICATION DATA: US 07/201,762
FILING DATE: 02-JUN-1988
APPLICATION DATA: US 07/180,548
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: MIT-4641AAA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1805 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-853-913-2

Alignment Scores:
Pred. No.: 6,07e-05 Length: 1805
Score: 162.50 Matches: 169
Percent Similarity: 34.81% Conservative: 113
Best Local Similarity: 20.86% Mismatches: 273
Query Match: 3.47% Indels: 255
DB: 1 Gaps: 40

US-09-787-016A-1 (1-2610) x US-07-853-913-2 (1-1805)
QY 3 CGCTGGCGCCGCGCCACTCCGCGGGGCTTCGGGAAATGGCTCGAGACCTAGAGGCCT 62
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
382 ArgThrProThrLeuAlaSerThrProIleProPheSerGluAlaProCysProPro 401
QY 63 ---CGGAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGAGTTGTTGAAAATACGAA 119
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
402 AsnAlaGluValArgAlaGluGluValProLeuSerLeuLeuGlnThrGlnAlaProGlu 421
QY 120 GCCTGTATAC---TCGTGAACAGTGGCTGACACAGAGTGTCTTCTG----- 161
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
422 ProLeuTrpLeuAlaThrValProSerSerSerAlaIleLeuProGluLeuGluGlu 441
QY 162 -----AGCTTGGCTGTCTGCTTG-----AGCTTGGCTGTCTGCTTG 179
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
442 ProGlyGlyLysGlnGlnGlyHisPheProAspLeuThrSerLeuAlaThrAsnLeu 461
QY 180 GACCCA-----GAGTTTCGTCTGCCAGGGTTTTT 209
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
462 AsnProHisHisProThrLeuGluAlaLysAspGlyGlySerSerGluSerArgValSer 481
QY 210 GGTGTGATTT----- 219
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
482 Ser-IlePheGlnGluAspGluGlyGlnIleTrpGluLeuValGluLysGluAlaAsp11 501
QY 220 -----AGATTTCAGGAAAGTGTCCAGCTTTCAGTGTTCGAGCAGGTATGACGA 272
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
501 eGluValLysValGluAsnSerSerAlaGlnLys---ThrGlnGluSerGlyLeuAsp-- 519
QY 273 CAAGGGCAGCCGAGCAATGAGGAGGACCATTAAGGCCATCAAAACCCACGAGCAAGAGTT 332
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

520 -----ThrGluGluThrGluAspSerGlnGlyProLeuGlnLysGluTh 534
QY CAGGAAACATGGGT-----TTTCGAAGGACCACTATCGCAACGAGAGGGCGC 383
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
534 rleuylsalaleuGlyGluGluProLeuMetSerLeuLysileGlnAsnTyrrGluThrAl 554
QY AGGG-----GAGCGGAGGCTGACCCACCTCGAGCGCCGACCC 419
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
554 aglyLysGluAsnCysAsnSerSerThrGlyHisLeuGlyThrLeuGluGlyProG1 574
QY -----TCCTCTGGGCGCAGT----- 456
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
574 uLysGluLysGlnIleProLeuLysSerLeuGluGluLysAsnValGluSerGluLysTh 594
QY -----GGGAGGCGAGCCCAACGCGACTGA 479
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
594 rleuGluAsnGlyValProValLeuSerGluLeuLysGluAspThrArgThrG1 614
QY -----CGCTGGAGCAGTTCCTGACCATTCGCGCG-- 510
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
614 uAspGlnGluLeuMetSerProLysGlyThrLeuLysArgPheSerSerLeuLysG1 634
QY -----CGCGCGCGGCG--AGGACATCGCTGTCTCCCT 542
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
634 uSerGlnGluValValArgProSerLysGluGlyAsnLeuGluSerTrpThrAlaPheL 654
QY -----CGCGCGCGGCG--AGGACATCGCTGTCTCCCT 542
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
543 CGAGGATTCGTGAGCCCGCTCTCGCCCGCCACAGACGCGGAGACAGCCCTCGAGG 602
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
654 sGluGluSerGlnHisProLeuGlyPheProGlyAlaGlu---AspGlnMetLeuGluAr 673
QY CAGCGTGGAAAGCGCTTCTGAGACCAAGAGCGGCCCGCAGTCTGCTTCACAGCTGTGAA 662
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
673 gLeuValGlu-----LysGluAspGlnSerPheProArgSerProGluGluGluAspG1 691
QY -----CGACCAGCGCTCTTCTGAAAGGCGGAGGAGGATCCAGAGTCCGCTCGGAGAA 836
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
721 sGluSerGlnGluSerLeuArgSerProGluGluGluAspGlnGluAlaGlyArgSerLe 741
QY CGCGCGGAGGAGGCTCCGCGCGAGCTGTGGC---TCCGAGGCGCAGTGACACTGTGGA 893
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
741 uGlnLysGluAsn-----GlnGluProLeuGlyTyrrGluGluAlaGluAspGlnMetLe 759
QY GGGCTCTCTGCCAGTAAGACGAGCGCCGAGAACGATCAGCGGGTGTGTCTCC----- 946
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
759 uGluArgLeuIleGluLysGluSerGlnGlu-SerLeuLysSer-----ProGluGlu 777
QY -----AGGCTGGGAAGATGACAGAGAGAGTAAAGTTGGAGGGAAGAGCGG-- 991
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
777 sGlnArgIleGlyLysProLeuGluArgGluAsnGlnArgProLeu----- 813
QY -----CTCAGGACATCAAGATGAGAGCGCTCGAGACTTGGG 1028
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
992 -----CTCAGGACATCAAGATGAGAGCGCTCGAGACTTGGG 1028
QY -----CTCAGGACATCAAGATGAGAGCGCTCGAGACTTGGG 1028
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
797 lAsnGlnGluThrPheValProLeuGluSerArgAsnGlnArgProLeu----- 813
QY -----CTCAGGACATCAAGATGAGAGCGCTCGAGACTTGGG 1028
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1029 CCGACCGAGGCTGATGATGAGGTTTACACCCCAAGCCCTGATTGTCATTTCGCGCCA 1088
QY -----ArgSerLeuGluValGluGlu-GluGluGlnArgIleValLysProLeuLys 831
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
814 -----ArgSerLeuGluValGluGlu-GluGluGlnArgIleValLysProLeuLys 831
QY GCTCACAACAACAGGTTTATGATTG-----CTGTGACCGCTGTGGAAGATG 1136
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
832 ValSerGlnAspSerLeuGlySerLeuAlaGluGluAsnValGlnProLeuArgTyrrLeu 851
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1137 GTTTCATGCGGATTGTGTGGCGATTCT-----GAGCCTCGAG 1175
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
852 Glu-GluAspAspCysIleAsnLysSerLeuLeuGluAspLysThrHisLysSerLeuG1 871
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```







Db 7325 -----ThrProAlaProGluAlaProAlaGlu 7333  
 QY 556 GAGCCACGCTCTGCCCGCC----- 576  
 Db 7334 GlnProLysProAlaProAlaProGlnProAlaProAlaProLysProAla 7353  
 QY 577 -----ACAGCGCGAGACAGCTCCGAGCGGACGCGTGAAGC 615  
 Db 7354 GluGlnProLysAlaGluLysThrAspGlnGlnAlaGluGluAspTyrAlaArg 7373  
 QY 616 GCTCTCTAG-----ACCAGAGCGCGCCCTCTCTCTCTCCAGCTGTG 660  
 Db 7374 SerGluGluGluTyrAsnArgLeuThrGlnGlnProProLysAlaGluLysProAla 7393  
 QY 661 -----AAGAACACACGCTCTCTGAAGAGTGAAGGAGGATGACCAC 708  
 Db 7394 ProAlaProLysProGluGlnProAlaProAlaProLysAsnSerLysGlyGluAla 7413  
 QY 709 GATGACACCTCCGATAGTACACGCTGCTGCTTGAAGAGCTTCAGAACTCCGCTT 768  
 Db 7414 GluGlnTyrArgSerAlaAlaGlyGlyAspLeuAlaAlaLysGlnValGluLeuGluLys 7433  
 QY 769 CGC-----AGGAAGCGGGAACAGGAGCCCTCTGAGGCGCCCTCTGAAGGATC 816  
 Db 7434 ThrGluAlaAspLeuLysLysAlaValAsnGluPro-----GluLysProAlaProAla--p 7452  
 QY 817 CAGAGTCGCTCGGAGAGCGCGGAGAGGCTCCCGCCGAGACTGTGGGCTCCGAG 876  
 Db 7452 xGluThrProAlaProGluAlaProAla--GluGlnProLysProAlaProAlaProGln 7471  
 QY 877 GCACGTGACACTGTGAGCGGCTCTGCCAGTACAGAGCGCCGAGAACGATCAGGG 936  
 Db 7472 ProAlaProAlaProLysProGluLysProAlaGluGln-----ProLysAlaGluLysPro 7490  
 QY 937 GTTGTCTCCAGCTGGAAGAT---GACAGAGAGTAAGTTGAGGGGAAG----- 987  
 Db 7491 AlaAspGlnGlnAlaGluGluAspTyrAspArgSerGluGluGluGluTyrAsnArgLeu 7510  
 QY 988 ---CGCGCTCAGGACATCAAGATGAGGCGCTGGAGACTTGGCGCGACCGAGCTCGAA 1044  
 Db 7511 ThrGlnGlnGlnProProLysAlaGluLysProAlaPro-----AlaProGlnProGlu 7528  
 QY 1045 TGTAGGTTACAGCCCGGCGCTCTGTATTTGCTATTTGCGCGCGCTCACAACACAGG 1104  
 Db 7529 GlnProAlaProAlaProLysSerLeuLysGluLeuAspGluSerAspSerGlu----- 7546  
 QY 1105 TTTATGATTTGCTGACCGCTCTGAAGATGTTTCATCGCGATTTGTGGGCAATTCT 1164  
 Db 7547 -----AspTyrValLysGluGlyPheArgAlaProLeuGlnSerGluLeu 7561  
 QY 1165 GAGGCTCGAGGAGGCTTTTGAAGGAATGGGAA-----GACTATATCTGCCCAAC 1218  
 Db 7562 AspAlaLysGlnAlaLysLeuSerLysLeuGluGluLeuSerAspLysIle----- 7578  
 QY 1219 TGCACCATTTGCAAGTGCAGGAGTACATCTCAGAACCGCA-----GAT 1266  
 Db 7579 -----AspGluLeuAspAlaGluLeuAlaLysLeuGluLysAsp 7591  
 QY 1267 CAGCAGGAAGCTAAATGAGACCTCGAGATGCTGATGGCAGCATTTGTACAGTATAGGA 1326  
 Db 7592 ValGluAspPheLys\*\*\*SerAspGlyGluGlnAlaGly-----GlnTyrLeuAla 7608  
 QY 1327 ACAATAGCAG-----AGTCTAGGAGACACCAAGGATAAGGCTAGATT 1374  
 Db 7609 AlaAlaGluGluAspLeuLeuAlaLysLeuAlaGluLeuGlnThrGluAlaAspLeu 7628  
 QY 1375 GAGAAAGCTCAATCCAAAGTGGGCAAGAACTCAAGATCTTC-----CAGCCTGTG 1428  
 Db 7629 LysLysAlaValAsnGluProGlyLysProAlaProAlaProAlaProGluThrProAla 7648  
 QY 1429 ATAGAGCGCTGTGCTCAAAATGATTGGGCCCGGCTGTCTACGTGGCG----- 1482  
 Db 7649 ProGluAlaProAlaGluGlnProLysProAlaProGluThrProAlaProAlaProLys 7668

QY 1483 -----CAGCCCGACTCGTCTACTGCACTAAT 1509  
 Db 7669 ProGluLysProAlaGluGlnProLysProGluLysProAlaAspGlnGlnAlaGluGlu 7688  
 QY 1510 GACTGTATCTCAACACGCGCGAGACATGAAGTTTCTAAGCTCAGGTAAGAAGACAG 1569  
 Db 7689 Asp-----TyrAlaArgSerGluGluGluTyrAsnArgLeuThrGlnGln 7704  
 QY 1570 AAGCCAAAGCTAAAGAAAGATGAAGT-----AAGCCAGAGAGAGCCAGCTCTCCG 1623  
 Db 7705 GlnProAlaProAlaGlnLysProGlnProAlaLysProGluLysProAlaGluGlu 7724  
 QY 1624 AAATCGGCTGCTCAGCGAGGTATTAAATCTCTCTGTGAC----- 1665  
 Db 7725 ProThrGlnProGluLysAspAlaGluLeuAlaLysLeuGluLysAsnValGluTyrPhe 7744  
 QY 1665 ----- 1665  
 Db 7745 LysLysThrAspAlaGluGlnThrGluGlnTyrLeuAlaAlaGluLysAspLeuAla 7764  
 QY 1666 ---AAGAGACCGAGCTCCAGAAAAAAGAGAGACACAGTGAAGAGAGCGTGTGCTCCT 1722  
 Db 7765 AspLysLysAlaGluLeuGluLysThrGluAlaAspLeuLysLysAlaValAsnGluPro 7784  
 QY 1723 GCGCGAGTGAAGCACTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1782  
 Db 7785 -----GluLysProAlaGluGluThrProAlaProAlaProLysProGluGlnPro 7801  
 QY 1783 AGCGATCACAAATCAATGCTAGTAAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1836  
 Db 7802 AlaGluGlnProLysProAlaProAlaProGlnProAlaProAlaProLysPro 7819  
 RESULT 12  
 US-10-164-595-56  
 ; Sequence 56, Application US/10164595  
 ; Patent No 6657054  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Origene Technologies, Inc  
 ; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides  
 ; FILE REFERENCE: IU 103 R1  
 ; CURRENT APPLICATION NUMBER: US/10/164,595  
 ; CURRENT FILING DATE: 2002-06-10  
 ; NUMBER OF SEQ ID NOS: 80  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 56  
 ; LENGTH: 779  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-10-164-595-56  
 Alignment Scores:  
 Pred. No.: 8,98e-05 Length: 779  
 Score: 158.50 Matches: 106  
 Percent Similarity: 32.67% Conservative: 59  
 Best Local Similarity: 20.99% Mismatches: 195  
 Query Match: 3.38% Indels: 145  
 Gaps: 19  
 DB:  
 US-09-787-016A-1 (1-2610) x US-10-164-595-56 (1-779)  
 QY 268 GAGCAGAAAGCGGACCGCGAGCAATGAGGAGGACCTTAAGCCCATCAAAACCCAGCAAA 327  
 Db 271 GluAspLysArgAspLeuIleSerArgGluIleSerLys----- 283  
 QY 328 GAGTTCAAGAAACATGGGGTTTTCGAGGAGCACATATCGCCAGCGAGAGGCGCAGGG 387  
 Db 284 -----PheArgAspThrHisLysLysLeuGluGluLysGly 296  
 QY 388 GAGCGGAGCTGACCCACTGCGGCGCGCCACCCACAGCAGCAGCTGCGCTGTCTCCTG 447  
 Db 297 LysLysGluLysGlu----- 301

QY 448 CCGCGCAGTGGGAGGAGCCAGCCAGCTGAGCGGTGGAGAGTTCTGACCAATTGG 507  
 Db 302 -----ArgGlnGluLeuGluysGluArgGluArgGluArgGluArgGlu 317  
 QY 508 CCGCGCGCGGAGGAGGAGTGCCTGCTCCTCGAGGATTCTGTCAGCCAGTCC 567  
 Db 318 ArgGluArgGluArgG----- 323  
 QY 568 TGCCCGCCAGCAGCCGAGAGCAGCTCCGAGGGCAGCGTGGAAAGCCCTTCGAGACC 627  
 Db 324 -----GluArgGluArgGluArgGluArgGluArgGluArgGluArgGluArg 339  
 QY 628 AGAGCGCGCCCGAGTCTGCTCCAGAGTGTGAGAGAGCAGCCCTCTTCGAARAG 687  
 Db 340 GluArgGluArgGluArgGluArgGluArgGluArgGluArgGluArgGluArg 359  
 QY 688 GTCAAGAGGAGGAGGATGACACGATGACCTCCGATGATGACAGCGATGCGCTGACCTTG 747  
 Db 360 GluArgGluArgGluArgGluArgGluArgGluArgGluArgGluArgGluArg 379  
 QY 748 AAGAGCTTCAGAAATCCCTCCAGAGAGCGGAGCAGAGCCCACTGAGAGGCCCTG 807  
 Db 380 ArgGluLeuGluArgGluysLeuArgGluys--GluAlaAlaTyrgInGluArg--Leu 397  
 QY 808 AAGAGGATCCAGATGCGCTCCGCGAAGAG-----CGCCGG 843  
 Db 398 LysAsnTrpGluLeuArgGluArgGluysThrArgGluTyrgInGluysGluAlaGluArg 417  
 QY 844 GAGAGGCTCCCGCAGAGCTGTGGCTCCGAGCC-----AGTGACACTGTGGAG 894  
 Db 418 GluGluGluArgArgGluMetAlaLysAlaLysArgLeuysGluPheLeuGlu 437  
 QY 895 GCGCTCTGCCAGTACAGGAGGAGCCGAGACGATCAGGGGTGTGTCAGGCTGG 954  
 Db 438 AspTyraAspAspArgAspProLysTyrrArgGlySerAlaLeuGluLysArg 457  
 QY 955 AAGAGTACAGAGAGTAAGTTGGAGGAGAGCGGCTCAGCACATCAAGATGAGGAG 1014  
 Db 458 LeuArgAspArgGluysGluMetGluAla-----AspGluArg 470  
 QY 1015 CTGGAGACTTGGCGCAGCGAGCCTGATGTGGGTTACGCCCAAGCGCTGTAT 1074  
 Db 471 -----AspArgLysArgGluysGluLeuGlu 480  
 QY 1075 TGCATTGCGCCAGCCTCAACAACAGGTTTATGATTGCTGTGACCGCTGTGAAGAA 1134  
 Db 480 ----- 480  
 QY 1135 TGCTTCATGCGGATTGTGTGGCATTTCTGAGCTCGAGGAGGCTTTTGAAGGAT 1194  
 Db 481 -----GluLeuArgGlnArgLeuLeuAlaGluGly 490  
 QY 1195 GGGAGAGCTATATCTGCCAAACTGCACCATTTCTGCAAGTCAGGATGAGACTTCATCA 1254  
 Db 491 HisProAsp-----ProAspAlaGluLeuGlnArgMetGluGlnGlu-----Ala 505  
 QY 1255 GAAAGCGGAGATCAGAGGAGTAATGAGAGCTCGAGATGCTGTATGCGCAGGATTGT 1314  
 Db 506 GluArgArgArgGlnProGlnLeuLeuGlnGlnProGluSerGluGluGlu----- 523  
 QY 1315 ACAAGTATAGCAACAATAGTACGAGTCTAGCAAGACCAAGGAGTAAAGGGTAGAATT 1374  
 Db 524 -----GluGluysGlnGluysGluGluysArgGluGluProMet 537  
 QY 1375 GAGAAAGCTGCMAATCCAGTGGCAAGAAATCAAG---ATCTTCAGCTGTGATA 1431  
 Db 538 GluGluGluGluGluProGluGlnLysProCysLeuysProThrLeuArgProLysSer 557  
 QY 1432 GAGCGCTGTGCTGAAATGATTATGCGCCCGCTGCTGTCTACGTGGCGACCGGAC 1491  
 Db 558 SerAlaProSerValSerAlaSerGlyAsn-----AlaThrProAsn 572  
 QY 1492 TCGGTGTACTGAGTAATGACTGT-----ATCTTCAAACACGCGGAGGAGCAATGAAG 1545

Db 573 ThrProGlyAspGluSerProCysGlyLeuLeuLeuLeuProHisGlu----- 587  
 QY 1546 TTCTTAAGCTCAGGTAAAGAACAGACCA-----AAGCCTAAAGAAAAGATGAAG 1596  
 Db 588 -----AsnSerProAspGlnGlnProGluGlnHisArgProLysIleGlyLeuSer 605  
 QY 1597 ATGAG-----CCAGAGAGCCAGTCTCCGAAATCGGTGCTCAGCCAGGTATT 1647  
 Db 606 LeuLysLeuGlyAlaSerAsnSerProGlyGlnProAsnSerValLysArgLysLeu 625  
 QY 1648 AAAATCTCTTCTGTGCACAGAGCCAGCTCCAGAAAAGAGACACAGTGAAGAG 1707  
 Db 626 ProValAspSerValPheAsnLysPheGluAspGluAspSerAspValProArgLys 645  
 QY 1708 GCAGTGGTGGTCCCT 1722  
 Db 646 ArgLysLeuValPro 650  
 RESULT 13  
 US-08-439-818A-5  
 ; Sequence 5, Application US/08439818A  
 ; Patent No. 5654145  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Fukuda, Michiko N.  
 ; TITLE OF INVENTION: Trophinin and Trophinin-Assisting  
 ; TITLE OF INVENTION: Proteins  
 ; NUMBER OF SEQUENCES: 22  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESS: Campbell and Flores  
 ; STREET: 4370 La Jolla Village Drive, Suite 700  
 ; CITY: San Diego  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 92122  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent in Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/439,818A  
 ; FILING DATE: 12-MAY-1995  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/317,522  
 ; FILING DATE: 04-OCT-1994  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Campbell, Cathryn A.  
 ; REGISTRATION NUMBER: 31,815  
 ; REFERENCE/DOCKET NUMBER: P-LA 1563  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (619) 535-9001  
 ; TELEFAX: (619) 535-8949  
 ; INFORMATION FOR SEQ ID NO: 5:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 778 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-439-818A-5  
 Alignment Scores:  
 Pred. No.: 0.000132 Length: 778  
 Score: 156.50 Matches: 170  
 Percent Similarity: 30.53% Conservative: 73  
 Best Local Similarity: 23.36% Mismatches: 272  
 Query Match: 3.34% Indels: 282  
 DB: 1 Gaps: 38  
 US-09-787-016A-1 (1-2610) x US-08-439-818A-5 (1-778)  
 QY 302 CTAGGCCATCA-----AACCCACCAAGAGGTTCAAGGA 337



QY 2086 GGTGAACGTTACCTTAATGATTCTTTAACTCTCTGTTTCACTCTCAGGCTCTGCT 2145  
Db 719 uThraIleHis-----CysPheHisgluAlaArgLeuAs 731  
QY 2146 AAGTATTGTTATCTCTCTATCCAGTCTGATGCTAGCCACCTGCGCGGACGCCAC 2205  
Db 731 pAspGluCysAlaPheTy-----ThrSerArgAlaSerPr 743  
QY 2206 ATCCACCCCT-----GTCTGCACATGAGTGTCTTGACA 2239  
Db 743 oSerGlyProThrArgValCysThrAsnProValAlaThr 756

RESULT 14

US-08-751-965-5  
Sequence 5, Application US/08751965  
Patent No. 5858360  
GENERAL INFORMATION:  
APPLICANT: Fukuda, Michiko N.  
TITLE OF INVENTION: Trophinin and Trophinin-Assisting  
TITLE OF INVENTION: Proteins  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell and Flores  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/751,965  
FILING DATE: Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/439,818  
FILING DATE: 12-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-LA 2252  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 778 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-751-965-5

Alignment Scores:  
Pred. No.: 0.000132 Length: 778  
Score: 156.50 Matches: 170  
Percent Similarity: 30.53% Conservative: 73  
Best Local Similarity: 21.36% Mismatches: 272  
Query Match: 3.34% Indels: 282  
DB: 2 Gaps: 38

US-09-787-016A-1 (1-2610) x US-08-751-965-5 (1-778)

QY 302 CTAGGGCTCA-----AACCCACAGCAAGAGTTCAGGA 337  
Db 95 LeuArgProSerProArgGlyGlnAsnValGlyProGlyProProAlaGlnThrGluAla 114  
QY 338 AACATGGGGTTTCGAGGAGCCACTATCGCCAGCGAGAGGCG----- 382  
Db 115 Pro---GlyThrIleGluPheValAlaAspProAlaAlaLeuAlaThrIleLeuSerGly 133

QY 383 CAGGGACGCGGAGGCTGACCCACTGGAGCGCCACCCACAGCAGCAGCTGGGCTCTGT 442  
Db 134 GluGlyValLysSer---CysHisLeuGlyArgGlnProSerLeuAlaLysArgValLeu 152  
QY 443 CCCTGCGCGCAGTGGGAGGAGC----- 466  
Db 153 ValArgGlySerGlnGlyThrThrGlnArgValGlnGlyValArgAlaSerAlaTyr 172  
QY 467 -----CCAGCGCAGCTGAGCGCGCTGGAGCAGT 493  
Db 173 LeuAlaProArgThrProThrHisArgLeuAspProAlaArgAlaSerCysPheSerArg 192  
QY 494 TCCTGACCATTTGCGCGCGCGCGGAGGAGCATGCTCTCTCTCCCTGGAGGAGT--- 550  
Db 193 -----LeuGluGlyProGlyProArgGlyArgThrLeuCysProGlnArgLeuGln 209  
QY 551 ---CTGCTGAGCGCCAGCTCTGCGCGCGCCACAGACCGCGAGACA-----GCTCCG 598  
Db 210 AlaLeuIleSerPro--SerGlyProSerPheHisProSerThrHisProSerPheGlnG 229  
QY 599 AGGCGAGCGTGGAAAGCGCTTCTGAGACCGAGAGCGCGCCCGCAGTCTGCTTCCACAGCTG 658  
Db 229 LuLeuArgArgGluThrAlaGlySerSerArgThrSerValSerGlnAlaSerGlyLeuL 249  
QY 659 TGAAGGAACGACCGCTCTCTGAAAGGTGAAAGGAGGAGGATGACCCAGATGACACCT 718  
Db 249 euLeuGluThrProValGlnProAlaPheSerLeuProLysGlyGlnArgGluValVal 269  
QY 719 CCGATAGTGACAGCGCATGCGCTGACCTTGAAGAGCTTCAGAAATCGC-----CTTCGCA 772  
Db 269 hrHisSerAspGluGlyValAlaSerLeuGlyLeuAlaGlnArgValProLeuArgG 289  
QY 773 GGAAGCGGAAACAGGAG----- 789  
Db 289 LuAsnArgGluMetSerHisThrArgAspSerHisAspSerHisLeuMetProSerProA 309  
QY 790 --CCCACTGAGAGCGCCCTGAAAGGATCCAGAGTCGCTGCGGAAGAGCGCGGAGG 847  
Db 309 laProValAlaGlnProLeu--ProGlyHisValValProCysProSerProPheGlyArg 328  
QY 848 -----AGGCTCCCGCGAGACTGTGGGCTCCGAGCGCAGTGCACACTGTGGAGGGCTCC 901  
Db 329 AlaGlnArgValPro-----SerProGlyProProThrLeuThrSertyrSer 344  
QY 902 -----TGCCCGAGTAAGCAGGACCGCG 922  
Db 345 ValLeuArgArgLeuThrValGlnProLysThrArgPheThrProMetProSerThrPro 364  
QY 923 AGAAGC-----ATCAGGGGGTTGTGTCCAGGCTGGGAAAGATGACAGAG 967  
Db 365 ArgValGlnGlnAlaGlnTrpLeuArgGlyValSerProGlnSerCysSerGluAspPro 384  
QY 968 AGAGTAAGTTGGAGGAAAGCGGCTCAG-----ACATCAAGATGAGGAGC 1015  
Db 385 AlaLeuProTrpGluGlnValAlaValArgLeuPheAspGlnGluSerCysIleArgSer 404  
QY 1016 CTGAGGACTTGGCGCGCAGCGAGCCTGGAATGTGAGGGTTACGACCCCA----- 1063  
Db 405 LeuGlu-----GlySerGlyLysProProValAlaThrProSerGlyProHis 420  
QY 1064 -----ACGCCCTGTATTGCAATTTGCGCGCAGCTCACAACACAGGTTTATGATT 1114  
Db 421 SerAsnArgThrPro----- 425  
QY 1115 GCTGTGACCGCTGTGGAAGAAATGTTTCATGCGGATTTGTGGGCAATTTCTGAGGCTCGAG 1174  
Db 426 SerLeuGlnGluValLysIleGlnArgIleGlyIleLeuGlnGlnLeuLeuArgGlnGlu 445  
QY 1175 GGAGGCTTTTGGAAAGAAATGGGAGAGACTATATCTGCCCAACTGCACCATTTCTGCAAG 1234  
Db 446 -----ValGluGlyLeuValGlyGlyGlnCysValPro----- 456

QY 1235 TGCAGGATGAGACTCATTCAAAACGAGATCAGCAGGAGGAGCTAAATGAGACCTGGAG 1294  
Db 457 - - - - - LeuAsnGlyGlySer 462  
QY 1295 ATGCTGATGCCACCGATTGTACAGTATAGACAATAGACGAGAGTCTAGCGAAGACC 1354  
Db 463 LeuAspMet - - - - - 465  
QY 1355 AAGGGATAAGGTAGATTGAGAAAGCTGCAATCAAGTGGCAAGA - - - - - 1402  
Db 466 - - - - - ValGluLeuGlnProLeuLeuThrGluLeuSerArgThrLeuAsnAla 481  
QY 1403 - - - - - AGAACTCAAGATCTTCAGCTGTG - - - - - ATAGAGCGCTGTGCTGCTCA 1449  
Db 482 ThrGluHisAsnSerGlyThrSerHisLeuProGlyLeuLeuHisSerGlyLeuPro 501  
QY 1450 AAA - - - - - TGTATTGGCCCGGTGTGTCACTGGCGCAG - - - - - 1485  
Db 502 LysProCysLeuProGluGluCysGlyGluProGlnProCysProAlaGluProGly 521  
QY 1486 CCGGACTCGGTGACTGACGATAGCTATGCTCTCAACACCGCCGAGCGACAATGAG 1545  
Db 522 ProProGluAlaPheCysArgSerGluProGluLeuProGluProSerLeuGlnGluGln 541  
QY 1546 TTCTAAGCTCAGTAAAGAACAGACGAAAGCCTAAAGAAAGATGAAGTGAAGCCA 1605  
Db 542 Leu - - - - - GluValProGluProTyProProAlaGluProArgPro 555  
QY 1606 - - - - - GAGAACCCAGTCTTCGAAATCGGTCTCAGGAGGATTT 1647  
Db 556 LeuGluSerCysCysArgSerGluProGluLeuProGluSerArgGlnGluGlnLeu 575  
QY 1648 AAAATCTCTCTGTGCACAGACAGACGCTCCAGAAAAAAGAGACCAAGTGAAGAG 1707  
Db 576 GluVal - - - - - ProGluProCysPro - - - - - 582  
QY 1708 GCAGTGTGTGTCCTGCGCGAGTGAAGCTCGGGAAGAGCAGCTTGTGAGAGCAGC 1767  
Db 583 - - - - - ProAlaGluProArgProLeuGluSerTyCysArgIleGluProGlu 598  
QY 1768 ACCCGTGTGGGGAGCGATCAATACATGACGTAAAGCAGAAAGATGCTGCT 1827  
Db 599 IleProGluSerArgGlnGlnGlnLeuGluValProGluProCysProProAlaGlu 618  
QY 1828 CCTCGCGCTCACTGTTGTATAATGTATGATCACCCTAGGGGTGCGCTCTCGGACCCC 1887  
Db 619 ProGlyPro - - - - - LeuGlnPro 624  
QY 1888 TCCGTTCTTCTGATAGCATCCCTCGGCGCTGTCCA - - - - - GGACTGGGAGTTGCA 1941  
Db 625 SerThrGlnGlnSerGlyProProGlyProCysProArgValGluLeuGlyAlaSer 644  
QY 1942 GCTTTGTTAAGCTGATCAGAG - - - - - 1965  
Db 645 GluProCysThrLeuGluHisArgSerLeuGluSerSerLeuProProCysCysSerGln 664  
QY 1966 - - - - - CACCGGCTGCACCATCAGCGGGAGAGCAGACCC 1998  
Db 665 TrpAlaProAlaThrSerLeuLeuPheSerSerGlnHisProLeuCysAlaSerPro 684  
QY 1999 - - - - - ATGTCAGATGCTCTCTGCTGCC - - - - - CTG 2025  
Db 685 ProIleCysSerLeuGlnSerLeuArgProProAlaGlyGlnAlaGlyLeuSerAsnLeu 704  
QY 2026 TGTCATCCCTAGTCTCTCAGGACTCTCTGCTCACTGTTTCCAAAGCTGTAAACCTCACT 2085  
Db 705 AlaProArgThr - - - - - LeuAlaLeuArgGlu-SerLeuLysSerCysLe 719  
QY 2086 GGTGAAGCTTCACTTAATGATTGATTCTTAACTCTGTTTCACTCTCAGGCTCTGTT 2145  
Db 719 uThrAlaIleHis - - - - - CysPheHisGluAlaArgLeuAs 731  
QY 2146 AAGTATTGTATTCTTCTTATCCAGTCTGATGTCATGATGATGATGATGATGATGATGAT 2205

Db 731 pAspGluCysAlaPheTy - - - - - ThrSerArgAlaSerPr 743  
QY 2206 ATCCACCCCT - - - - - GTCTGCACATGAGTTGTTCTGACA 2239  
Db 743 oSerGlyProThrArgValCysThrAsnProValAlaThr 756  
RESULT 15  
US-08-738-975-5  
; Sequence 5, Application US/08738975  
; Patent No. 5880267  
; GENERAL INFORMATION:  
; APPLICANT: Fukuda, Michiko N.  
; TITLE OF INVENTION: Trophinin and Trophinin-Assisting  
; TITLE OF INVENTION: Trophinin  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSES: Campbell and Flores  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/738,975  
; FILING DATE: herewith  
; CLASSIFICATION: 435  
; PRIOR APPLICATION NUMBER:  
; APPLICATION NUMBER: US 08/439,818  
; FILING DATE: 05-Dec-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-LA 2251  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 778 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-738-975-5  
Alignment Scores:  
Pred. No.: 0.000132 Length: 778  
Score: 156.50 Matches: 170  
Percent Similarity: 30.53% Conservative: 73  
Best Local Similarity: 21.36% Mismatches: 272  
Query Match: 3.34% Indels: 282  
DB: 2 Gaps: 38  
US-09-787-016A-1 (1-2610) x US-08-738-975-5 (1-778)  
QY 302 CTAAGGCCATCA - - - - - AACCCACGACCAAGAGTTCCAGGA 337  
Db 95 LeuArgProSerProArgGlyGlnSerValGlyProGlyProAlaGlnThrGluAla 114  
QY 338 AAACATGGGGTTTTCGAGGACCACTATCGCCAGCGAGAGGGCG - - - - - 382  
Db 115 Pro---GlyThrIleGluPheValAlaAspProAlaAlaLeuAlaThrIleLeuSerGly 133  
QY 383 CAGGGACCGCGGAGGCTGACCCACTGGAGCCGCCACCCACGACGAGCTGGGCTGT 442  
Db 134 GluGlyValLysSer---CysHisLeuGlyArgGlnProSerLeuAlaLysArgValLeu 152  
QY 443 CCTGTGGGCGCAGTGGGAGGAGC - - - - - 466





Search completed: April 28, 2004, 11:03:03  
Job time : 89.5 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 28, 2004, 10:33:22 ; Search time 119 Seconds  
(without alignments)  
12394.161 Million cell updates/sec

Title: US-09-787-016A-1  
Perfect score: 4689  
Sequence: 1 ctccgtggccgcgcgcac.....actcttaagatcatatcctg 2610

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-O=/cpn2.1/USPTO.spool\_p/US09787016/runat.28042004.094328.26896/app\_query.fasta\_1.2759  
-DB=A\_Geneseq.29Jan04 -CFMT=fastan -SUPPLX=rag -MINMATCH=0.1 -LOOPCL=0  
-LOOPT=0 -UNIT=bits -START=1 -END=1 -MATRIX=BLOSUM62 -TRANS=human+0.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM-ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09787016 @CNC 1.1 91 @runat.28042004.094328.26896 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG  
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A\_Geneseq.29Jan04:\*

- 1: Geneseq1980s:\*
- 2: Geneseq1990s:\*
- 3: Geneseq2000s:\*
- 4: Geneseq2001s:\*
- 5: Geneseq2002s:\*
- 6: Geneseq2003as:\*
- 7: Geneseq2003bs:\*
- 8: Geneseq2004s:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2989	63.7	562	3 AAY67579	Human dea
2	2981	63.6	562	4 AAB3638	Human pro
3	2432	51.9	1191	4 AAM80219	Human pro
4	2352.5	50.2	775	4 ABG22389	Novel hum
5	2311	49.3	647	4 AAM79235	Human pro
6	2154.5	45.9	614	3 AAY67580	Murine de
7	1551	33.1	330	4 ABG22388	Novel hum
8	958	20.4	181	3 AAB43724	Human can
9	898	19.2	167	4 ABG22387	Novel hum
10	494	10.5	108	5 ABP05496	Human ORF

11	332	7.1	2016	4 ABB71487	Abb71487 Drosophil
12	232	4.9	1539	4 AAM78908	Aam78908 Human pro
13	232	4.9	1728	4 AAM32334	Aam32334 Human pol
14	232	4.9	1766	4 AAM79892	Aam79892 Human pro
15	232	4.9	1766	4 AAM41020	Aam41020 Human pol
16	232	4.9	2039	6 ABR64258	Abrr64258 Angiogene
17	232	4.9	2039	7 ADD45832	Add45832 Human pro
18	217	4.6	663	4 ABB70523	Abb70523 Drosophil
19	212.5	4.5	386	4 ABB70522	Abb70522 Drosophil
20	212	4.5	3238	4 ABB71715	Abb71715 Drosophil
21	203.5	4.3	656	4 ABUS3239	Abus3239 Human tes
22	190.5	4.1	772	4 ABG22981	Abg22981 Novel hum
23	185	3.9	636	4 ABG03757	Abg03757 Novel hum
24	182.5	3.9	704	7 ADE60099	Ade60099 Rat prote
25	182.5	3.9	704	7 ADD47584	Add47584 Rat prote
26	182.5	3.9	704	7 ADE60734	Ade60734 Rat prote
27	182.5	3.9	704	7 ADE60097	Ade60097 Rat prote
28	182.5	3.9	704	7 ADD47582	Add47582 Rat prote
29	182.5	3.9	704	7 ADD47665	Add47665 Rat prote
30	182	3.9	699	4 AAM79156	Aam79156 Human pro
31	181.5	3.9	268	4 AAU16305	Aau16305 Human nov
32	181.5	3.9	268	6 ABUS5374	Abus5374 Human nov
33	181	3.9	2759	6 AAU16418	Aau16418 Human nuc
34	180.5	3.8	669	2 AAM37483	Aam37483 Mouse liv
35	180	3.8	718	4 AAM80140	Aam80140 Human pro
36	178	3.8	1938	6 ABE98398	Abbe98398 Streptomy
37	175.5	3.7	2004	5 ABG95113	Abg95113 Human tra
38	174.5	3.7	2289	2 AAM14987	Aam14987 Protein d
39	174	3.7	107	4 ABB16825	Abb16825 Human ner
40	173.5	3.6	1938	6 ABP76578	Abp76578 Streptomy
C 41	173.5	3.6	1938	6 ABP76579	Abp76579 Streptomy
C 42	172.5	3.6	1061	2 AAM87504	Aam87504 Human N-m
43	172	3.7	676	4 AAM47218	Aam47218 Human NOV
44	172	3.7	676	5 ABG05039	Abg05039 Human NOV
45	171.5	3.7	555	4 AAG75073	Aag75073 Human col

ALIGNMENTS

RESULT 1

AAY67579

ID AAY67579 standard; protein; 562 AA.

XX

AC AAY67579;

XX

DT 19-JUN-2000 (first entry)

XX

DE Human death inducer-obliterat 1 (DIO-1) polypeptide.

XX

KW Death inducer-obliterat 1; DIO-1; cell death; cancer; tumour; human; autoimmune disease; cytostatic; immunosuppressive; antidiabetic; antirheumatic; antiinflammatory; antiproliferative.

XX

OS Homo sapiens.

XX

EH Key

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

XX (CNSJ) CONSEJO SUPERIOR INVESTIGACIONES CIENTIF.  
PA (BANN/) BANNERMAN D G.  
XX Alonso CM, Domingo DG, Grandien A, Leonardo B, Martinez P;  
XX WPI; 2000-271426/23.  
DR N-PSDB; AA290578.  
XX New DNA encoding human and murine death inducer-obliterator 1  
PT polypeptides, useful in the treatment of cancer, autoimmune diseases,  
PT diabetes, rheumatoid arthritis, benign tumors, malignant tumors and  
PT hyperproliferative skin disorders.  
XX Claim 6; Fig 1C; 27pp; English.  
XX The invention provides nucleic acids encoding the human and murine death  
CC inducer-obliterator 1 (DIO-1) polypeptides. The polypeptides can be  
CC expressed by standard recombinant methodology. The DIO-1 polypeptides,  
CC agonists and antagonists are used as a medicament for treating diseases  
CC characterized by an alteration in cell death or by hyperproliferation,  
CC e.g. cancer, autoimmune diseases, diabetes, rheumatoid arthritis. Benign  
CC tumors, malignant tumors or hyperproliferative skin disorders. They are  
CC also useful in the treatment of metabolic, proliferative or inflammatory  
CC conditions. The present sequence represents the human DIO-1 polypeptide  
XX  
SQ Sequence 562 AA;  
  
Alignment Scores:  
Pred. No.: 2,82e-224 Length: 562  
Score: 2989.00 Matches: 562  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 63.74% Indels: 0  
DB: 3 Gaps: 0  
  
US-09-787-016A-1 (1-2610) x AAY67579 (1-562)  
  
QY 265 ATGAGACAGAAAGCGACCGAGCAATGAGAGGACCTAAGCCATCAAAACCCACGAG 324  
Db 1 MeAspAspLysGlyAspProSerAsnGluGluAlaProLysAlaLeuLysProThrSer 20  
  
QY 325 AARAGAGTTCAGGAAACATGGGGTTTTCGAGGACCACTATCCCAAGCGAGAGGCGCA 384  
Db 21 LysGluPheargLysThrTrpGlyPheArgArgThrThreAlaLysArgGluGlyAla 40  
  
QY 385 GGGAGCGGAGGCTGACCCACTCGAGCGCCACCCCGACGACGAGCTGGCGCTGTC 444  
Db 41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnGlnLeuGlyLeuSer 60  
  
QY 445 CTGGCGCGAGTGGGAGCGAGCCCAAGCGCACTGAGCGCGTGGAGCAGTTCCCTGACCAT 504  
Db 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThre 80  
  
QY 505 GCGCGCGCGCGGAGGAGAGTCCCTGTCCTCGAGAGATTCGTGAGCCGACG 564  
Db 81 AlaArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100  
  
QY 565 TCCTGCCCCCGACAGACGCGGAGACAGCTCCGAGGCGAGCGTGGAAAGCGCTTCGAG 624  
Db 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120  
  
QY 625 ACCAGAGCGGCGCCCGAGTCTGCTTCCACAGCTGTGAGGAGACGACGCTCTCTGAA 684  
Db 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140  
  
QY 685 AAGGTGAAGGAGGGATACCGACGATGACACTCCGATAGTCACGAGCGATGCGCTGACC 744  
Db 141 LysValLysGlyLysAspPheHisAspThrSerAspSerAspSerAspGlyLeuThr 160  
  
QY 745 TTGAAGAGGCTTCAGATCCGCTTCGAGGAGCGGAAACAGAGGCCCATGAGAGGCC 804  
Db 161 LeuLysGluLeuGlnAsnArgLeuArgArgLysArgGluGlnGluProThrGluArgPro 180

QY 805 CTGAAGGGATCCAGATCGCTCCGGAAGAACGCCGGGAGGAGGTCCCGCCGAGACT 864  
Db 181 LeuLysGlyIleGlnSerArgLeuArgLysLysArgGluGluGlyProAlaGluThr 200  
  
QY 865 GTGGCTCCGAGGCGCAGTGCACACTGTGGAGGCGCTCTGCCAGTAACAGACAGAGCCGAG 924  
Db 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGlu 220  
  
QY 925 AACGATCAGGGGGTGTGTCCTCCAGGCTGGGAAAGATCACAGAGAGTAAGTTGAGGGA 984  
Db 221 AsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGluGly 240  
  
QY 985 AAGCGGCTCAGGACATCAAGATGAGGAGCTCGAGACTTGGCGCCGACCAAGCCTCAA 1044  
Db 241 LysAlaAlaGlnAspIleLysAspGlnGluProGlyAspLeuGlyArgProLysProGlu 260  
  
QY 1045 TGTGAGGTTACGACCCCAAGCGCTGTATTTCATTTGCCGCCAGCTCACACAAACAGG 1104  
Db 261 CysGluGlyTyAspProAsnAlaLeuTyCysIleCysArgGlnProHisAsnAsnArg 280  
  
QY 1105 TTTATGATTTCTGTCGACCGCTGTGAGAAAGTGGTTTCATGGCGATTGTGGGCAATTCT 1164  
Db 281 PheMetIleCysCysAspArgCysGluGluTrpPheHisGlyAspCysValGlyIleSer 300  
  
QY 1165 GAGCTCGAGGAGGCTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCAACTGCACC 1224  
Db 301 GluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyIleCysProAsnCysThr 320  
  
QY 1225 ATTCTGCAAGTGCAGATGAGACTCATTTCAGAAAGCGGAGATCACAGAGAGCTAAATGG 1284  
Db 321 IleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLysTrp 340  
  
QY 1285 AGACCTCGAGATGCTGATGGCACCGATGTCACATATAGAAACAAATAGAGCAGAGTCT 1344  
Db 341 ArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSer 360  
  
QY 1345 AGCGAAGCAAGGGATAAAGGTAGAAATGAGAAGCTGCAAATTCGAAGTGGCAGAGAG 1404  
Db 361 SerGluAspGlnGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLysLys 380  
  
QY 1405 AAACCTCAGACTTCCAGCCTGTGTATAGAGCGCTGTGCTCAAAATGATTTGGCCCC 1464  
Db 381 LysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGlyPro 400  
  
QY 1465 GGGTCTGTCTCAGTGGCGCAGCCGACTCGGTGTACTCAGTAATGACTGTATCTCAAA 1524  
Db 401 GlyCysCysHisValAlaGlnProAspSerValTyCysSerAsnAspCysIleLeuLys 420  
  
QY 1525 CACGCGGAGGACAAATGAAGTTTCTAGCTCAGGTAAAGACAGACAGCAAGCCTAAA 1584  
Db 421 HisAlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysProLys 440  
  
QY 1585 GAAAAGATGAAGATGAAGCCAGAGAACCCAGCTCTCCGAAATGCGGTGCTCAGGCAGGT 1644  
Db 441 GluLysMetLysMetLysProGluLysProSerLeuProLysCysGlyAlaGlnAlaGly 460  
  
QY 1645 ATTAAATCTCTCTGTGACAGAGACAGCTCCAGAAAAAAGAGACACAGCAGTGAAG 1704  
Db 461 IleLysIleSerSerValHisLysArgProAlaProGluLysLysGluThrThrValLys 480  
  
QY 1705 AAGGCAAGTGGTGGTCCCTCGCGGAGTGAAGCACTCGGGAAGGAGACAGCTTGTGAGAGC 1764  
Db 481 LysAlaValValValProAlaArgSerGluAlaLeuGlyLysGluAlaCysGluSer 500  
  
QY 1765 AGCAGCGCGTGTGGGCGAGCGATCAATTCATGCGTAAAGCCAGAAAGACTGTCT 1824  
Db 501 SerThrProSerTrpAlaSerAspHisAsnTyAsnAlaValLysProGluLysThrAla 520  
  
QY 1825 GCTCCCTCGCTCCTGTTGTATAAATGATGTATCACTAGGGGTGGCCCTCTCGGAC 1884  
Db 521 AlaProSerProSerLeuLeuTyLysCysMetTyHisLeuGlyValGlyLeuLeuAsp 540

QY 1885 CCTCCCGTCTTCTTCTGATAGCCATCCCTGGGCTGTCAGGACTGGGAGTTGCAGCT 1944  
DB 541 ProSerArgSerPheTrpIleAlaIleProTrpAlaCysProGlyLeuGlyValAlaAla 560  
QY 1945 TTGTGT 1950  
DB 561 LeuCys 562  
RESULT 2  
ID AAB93638  
XX AAB93638 standard; protein; 562 AA.  
XX AAB93638;  
XX 26-JUN-2001 (first entry)  
XX Human protein sequence SEQ ID NO:13130.  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
XX Homo sapiens.  
XX EP1074617-A2.  
XX 07-FEB-2001.  
XX 28-JUL-2000; 2000BP-00116126.  
XX 29-JUL-1999; 99JP-00248036.  
XX 27-AUG-1999; 99JP-00300253.  
XX 11-JAN-2000; 2000JP-00118776.  
XX 02-MAY-2000; 2000JP-00183767.  
XX 09-JUN-2000; 2000JP-00241899.  
XX (HELI-) HELIX RES INST.  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.  
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.  
XX Claim 8; SEQ ID NO 13130; 2537pp + Sequence Listing; English.  
XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence and 3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention  
XX Sequence 562 AA;  
SQ

Alignment Scores:  
Pred. No.: 1.19e-223 Length: 562  
Score: 2981.00 Matches: 561  
Percent Similarity: 99.82% Conservative: 0  
Best Local Similarity: 99.82% Mismatches: 1  
Query Match: 63.57% Indels: 0  
DB: 4 Gaps: 0  
US-09-787-016A-1 (1-2610) x AAB93638 (1-562)  
QY 265 ATGGACGACAAAGCGGACCCGAGCAATGAGGAGGACCTAAGGCCATCAAAACCCACGAGC 324  
DB 1 MetAspAspLysGlyAspProSerAsnGluGluAlaProLysAlaIleLysProThrSer 20  
QY 325 AAAGAGTTTCAGGAAACATGGGGTTTTTCAGAGGACCACTATCGCCCAAGCGAGAGGGCGCA 384  
DB 21 LysGluPheArgLysThrTrpGlyPheArgArgThrThrIleAlaLysArgGluGlyAla 40  
QY 385 GGGGACGCGGAGGCTGACCCACTGGAGCGCGCCACCCACAGCAGCAGCTGGGCTGTGCC 444  
DB 41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnGlnLeuGlyLeuSer 60  
QY 445 CTGCGGCGCAGTGGGAGGACGCCCAAGCGCACTGAGCGCGTGGAGCAGTTCTCTGACCAT 504  
DB 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGluPheLeuThrIle 80  
QY 505 GCGCGCGCGCGGCGGAGGAGCATGCTGTCTCTCGCTGGAGGATCTCTGAGGCCCGCC 564  
DB 81 AlaArgArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100  
QY 565 TCTCTCCCGCGCCACAGAGCGCCGAGCAGCTCCGAGGGCAGCGTGGAAAGCCCTTCTGAG 624  
DB 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120  
QY 625 ACCGAGAGCGGCGCCCGAGTCTGCTTCCACAGCTGTGAAGGAAACGACGAGCTCTTCTGAA 684  
DB 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140  
QY 685 AAGGTGAAGGAGGGGATGACACCATGACACTCCGATAGTGACAGCAGCATGCCCTGACC 744  
DB 141 LysValLysGlyLysAspAspHisAspAspThrSerAspSerAspSerAspGlyLeuThr 160  
QY 745 TTGAAGAGCTTCAGAAATCGCTTCGAGGAAGCGGAAACAGGAGCCCACTGAGAGGCC 804  
DB 161 LeuLysGluLeuGlnAsnArgLeuArgLysArgLysGluGlnGluProThrGluArgPro 180  
QY 805 CTGAAGGATCCAGAGTCCGCTCGGAGAGAGCGCGGAGGAGGCTCCCGGAGACT 864  
DB 181 LeuLysGlyIleGlnSerArgLeuArgLysLysArgGluGluGluGluProAlaGluThr 200  
QY 865 GTGGGCTCCGAGCGCCAGTGACACTGTGAGGGCGCTCTGCCCCAGTAAGCAGGAGGCCGAG 924  
DB 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGlu 220  
QY 925 AACGATCAGGGGTGTGTCTCCAGGCTGGGAAAGATGACAGAGAGTAAGTTGGAGGCA 984  
DB 221 AsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGluGly 240  
QY 985 AAGGCGGCTCAGACATCAACATCAGAGAGCTGCGAGACTTGGCGCGACCCGAGCCTGAA 1044  
DB 241 LysAlaAlaGlnAspIleLysAspGluGluProGluAspLeuGlyArgProLysProGlu 260  
QY 1045 TTGAGGGTTACGAGCCCAACCGCTGTATTGCTATTTGGCGGAGTGTGTGGGCATTTCT 1164  
DB 261 CysGluGlyTyrAspProAsnAlaLeuTyrCysIleCysArgGlnProHisAsnAsnArg 280  
QY 1105 TTTATGATTTGCTGTGACCGCTGCAAGATGCTTTTCATGGCGATTTGTGTGGGCATTTCT 1164  
DB 281 PheMetIleCysCysAspArgCysGluGluTrpPheHisGlyAspCysValGlyIleSer 300  
QY 1165 GAGGCTCGAGGAGGCTTTTGAAGAGGATGGGAGAGACTATATCTGCCCAACTGCACC 1224  
DB 301 GluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyrIleCysProAsnCysThr 320



109	Db	rApSerAspSerAspGlyLeuThrLeuLysGluLeuGlnAsnArgLeuArgArgLysAr	129
780	QY	GGAAAGAGGCCACCTAGAGAGGCCCTGAAAGAGCATCCACAGCTCGCTCGGAAGAGCG	839
129	Db	gGluGlnGluProThrGluArgProLeuLysGlyIleGlnSerArgLeuArgLysLysAr	149
840	QY	CCGGAGAGGGTCCCGCCGAGACTGTGGCTCCGAGGCCAGTGACACTGTGAGGGCGT	899
149	Db	gArgGluGluGlyProAlaGluThrValGlySerGluAlaSerAspThrValGluGlyVa	169
900	QY	CTTGCCCACTAACAGAGGCCCGAGACAGTACAGGGGTGTCTCCACAGCTCGGAAGCA	959
169	Db	lLeuProSerLysGlnGluProGluAsnAspGlnGlyValValSerGlnAlaGlyLysAs	189
960	QY	TGACAGAGAGTAAGTTGGAGGGAAGCGGCTCAGGACATCAAAAGATGAGAGCGCTG	1019
189	Db	pAspArgGluSerLysLeuGluGlyLysAlaAlaGlnAspIleLysAspGluGluProG	209
1020	QY	AGACTTGGCCGACCGAGGCTGAATGTGAGGTTAGGACCCCAAGCCCTGTATTCGAT	1079
209	Db	yAspLeuGlyArgProLysProGluCysGluGlyTyArgProAsnAlaLeuTyCysIl	229
1080	QY	TTCCGCGCACCTCACAAACAGGTTTAAATTTGCTGTACCGCTGTCAAGAAATGGTT	1139
229	Db	eCysArgGlnProHisAsnAsnArgPheMetIleCysCysAspArgCysGluGluTrpPh	249
1140	QY	TCATGGCGATGTGTGGGCATTTCTGAGGCTCAGGAGAGGCTTTTGGAAAGGAATGGGA	1199
249	Db	eHisGlyAspCysValGlyIleSerGluAlaArgGlyArgLeuLeuGluArgAsnGlyGl	269
1200	QY	AGACTATCTGCCAAACTGCACCATTCCTGCAAGTGCAGATGAGACTCATTTCAAGAAC	1259
269	Db	uAspTyrlleCysProAsnCysThrIleLeuGlnValGlnAspGluThrHisSerGluTh	289
1260	QY	GGCAGATCAGCAGGAAGCTAAATCGAGACTCGAGATGTGTGTGGCACCCGATTTGTAAG	1319
289	Db	rAlaAspGlnGlnGluAlaIleYrTrpArgProGlyAspAlaAspGlyThrAspCysThrSe	309
1320	QY	TATAGGACATAGACGAGAGTCTACGAGACCAAGGATAAAGGTAGAATTGAGAA	1379
309	Db	rIleGlyThrIleGluGlnLysSerSerGluAspGlnGlyIleLysGlyArgIleGluLy	329
1380	QY	AGTGTCAAAATCCAGTGGCAAGAAACCTCAAGATCTTCCAGCCT-----	1425
329	Db	sAlaAlaAsnProSerGlyLysLysLysLeuLysIlePheGlnProGlyProVa	349
1425	QY	-----	1425
349	Db	lProThrGlnLeuProValLeuTrpGlnValLeuGluIleAlaValSerArgSerIleSe	369
1426	QY	-----GTGATAGAGCGCCTGGTGCTCAAA	1451
369	Db	rAlaPheThrLeuLeuHisCysIleSerCysLysValIleGluAlaProGlyAlaSerLy	389
1452	QY	ATGTATTGGCCCGGTGTGTCTACGTGGCGCACCCGACTCGGTGTACTGCAGTATGA	1511
389	Db	sCysIleGlyProGlyCysCysHisValAlaGlnProAspSerValTyCysSerAsnAs	409
1512	QY	CTGTATCTCTCAACACACCGCAGGCGCAATGAAGTTTCTAAGCTCAGCTAAAGAACAGAA	1571
409	Db	pCysIleLeuLysHisAlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLy	429
1572	QY	GCCAAAGCCTAAAGAAAGATGAAGTCAGCCAGAGAGCCCGACTCTTCGAAATCCGG	1631
429	Db	sProLysProLysGlnLysMetLysMetLysProGluLysProSerLeuProLysCysGl	449
1632	QY	TGCTCAGCAGGTATTAATCTCTCTGTGCACAGAGACCACTCCAGAAAAAAGA	1691
449	Db	yAlaGlnAlaGlyIleLysIleSerSerValHisLysArgProAlaProGluLysLysGl	469
1692	QY	GACCACAGTGAAGAGGCATGGTGTCTTCCTCGCGGAGTGAAGCATCTCGGAAGGAAGC	1751

Db	469	uThrThrValLysLysAlaValValProAlaArgSerGluAlaLeuGlyLysGluAla	1851
Qy	1752	AGCTTGTGAGCAGCAGCACGCCGCTGCTGGCGAGCGATCACAATTACAAATCGAGTAAGCC	1811
Db	489	aAlaCySgluSerSerThrProSerTrpAlaSerAspHisAsnTyzAsnAlaValLysPr	509
Qy	1812	AGAAAGACATGCTGCTCCCTCCGCGCTCACATGTTGTATAAA	1851
Db	509	oGluLysThrAlaAlaProSerProSerLeuLeuTyLys	522
RESULT	4		
ABG22389			
ID	ABG22389	standard; protein; 775 AA.	
XX	AC		
XX	AC	ABG22389;	
XX	XX		
DT	18-FEB-2002	(first entry)	
XX	XX		
DB	Novel human diagnostic protein #22380.		
XX	XX		
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;		
KW	food supplement; medical imaging; diagnostic; genetic disorder.		
XX	OS	Homo sapiens.	
XX	XX		
PN	WO200175067-A2.		
XX	XX		
PD	11-OCT-2001.		
XX	XX		
PF	30-MAR-2001; 2001WO-US0008631.		
XX	XX		
PR	31-MAR-2000; 2000US-00540217.		
PR	23-AUG-2000; 2000US-00649167.		
XX	XX		
FA	(HYSE-) HYSEQ INC.		
XX	XX		
PI	Drmanac RT, Liu C, Tang YT;		
XX	XX		
DR	WPI; 2001-639362/73.		
DR	N-PSDB; AAS86576.		
XX	XX		
PT	New isolated polynucleotide and encoded polypeptides, useful in		
PT	diagnostics, forensics, gene mapping, identification of mutations		
PT	responsible for genetic disorders or other traits and to assess		
PT	biodiversity.		
XX	XX		
PS	Claim 20; SEQ ID NO 52748; 103pp; English.		
XX	XX		
CC	The invention relates to isolated polynucleotide (I) and polypeptide (II)		
CC	sequences. (I) is useful as hybridisation probes, polymerase chain		
CC	reaction (PCR) primers, oligomers, and for chromosome and gene mapping,		
CC	and in recombinant production of (II). The polynucleotides are also used		
CC	in diagnostics as expressed sequence tags for identifying expressed		
CC	genes. (I) is useful in gene therapy techniques to restore normal		
CC	activity of (II) or to treat disease states involving (II). (II) is		
CC	useful for generating antibodies against it, detecting or quantitating a		
CC	polypeptide in tissue, as molecular weight markers and as a food		
CC	supplement. (II) and its binding partners are useful in medical imaging		
CC	of sites expressing (II). (I) and (II) are useful for treating disorders		
CC	involving aberrant protein expression or biological activities. The		
CC	polypeptide and polynucleotide sequences have applications in		
CC	diagnostics, forensics, gene mapping, identification of mutations		
CC	responsible for genetic disorders or other traits to assess biodiversity		
CC	and to produce other types of data and products dependent on DNA and		
CC	amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic		
CC	patent did not appear in the printed specification, but was obtained in		
CC	electronic format directly from WIPO at		
CC	ftp.wipo.int/pub/published_pt_sequences		
XX	XX		
SQ	Sequence 775 AA;		
	Alignment Scores:		



Prod. No.: 1,46e-174 Length: 775  
 Score: 2352.50 Matches: 485  
 Percent Similarity: 74.42% Conservative: 24  
 Best Local Similarity: 70.91% Mismatches: 58  
 Query Match: 50.17% Indels: 117  
 DB: 4 Gaps: 14

US-09-787-016A-1 (1-2610) x ABG22389 (1-775)

QY 202 GGGTTTTCGTTGCTATTAGGATTTTCAGGAAAGTGTCCAAAGCTTTCAGTCTTCGAGCA 261  
 DDb |||||  
 80 GlyGlyCysValAlaGlyLeuIleSerGlyLysSerValGlnAlaPheSerValGlyAla 99  
 QY |||||  
 261 ----- 261

QY 100 GlyArgTrpProLeuLeuArgValPheGlyGlyAsnValThrPheCysSerValValSer 119  
 DDb |||||  
 262 ----- GGTATGACGACAAAGCGCACCAAGCGCACCGAGCAAT 291  
 QY |||||  
 DDb 120 TrpSerLeuCysAlaTyrThrLeuProLeuGlyMetAspAspLysGlyAspProSerAsn 139  
 QY |||||  
 292 GAGGAGCACCTAAGGCCATCAAAACCACGACCAAGAGTTTCAGGAAACATCGGGTTT 351  
 DDb |||||  
 140 GluGluAlaProLysAlaIleLysProThrSerLysGluPheArgLysThrTrpGlyPhe 159

QY 352 CGAAGGACCACTATCCCAAGCGAGAGGCGCGAGGACCGCGAGGCTGACCCACTGGAG 411  
 DDb |||||  
 160 ArgThrThrIleAlaLysArgGluGlyAlaGlyAspAlaGluAlaAspProLeuGlu 179  
 QY |||||  
 412 CGCCACCCACGACGAGCTGGGCGCTGCTCCGCGGCGAGTGGGAGGAGCCCAAG 471  
 DDb |||||  
 180 ProProProGlnGlnGlnGlyLeuSerLeuArgSerGlyArgGlnProLys 199  
 QY |||||  
 472 CGCACTGAGCGCTGGAGCAGTTCTCTGACCATTCGCGCGCGCGCGAGGAGGAGCATG 531  
 DDb |||||  
 200 ArgThrGluArgValGluGlnPheLeuThrIleAlaArgArgGlyArgSerMet 219  
 QY |||||  
 532 CTTGTTCTCCGTGAGATTCGTGTAGCCACCTCTGCTGCCGCCACAGACCCGAGCA 591  
 DDb |||||  
 220 ProValSerLeuGluAspSerGlyGluProThrSerCysProAlaThrAspAlaGluThr 239  
 QY |||||  
 592 GCCTCCGAGGCGAGCTGGAAGCGTCTTGACACGAGAGCGCGCCCGCTGCTCTCC 651  
 DDb |||||  
 240 AlaSerGluGlySerValGluSerAlaSerGluThrArgSerGlyProGlnSerAlaSer 259  
 QY |||||  
 652 ACAGCTGTGAAGAACACGACGCTCTTCTGAAAGGTGAAGAGGGGATGACCAAGAT 711  
 DDb |||||  
 260 ThrAlaValLysGluArgProAlaSerSerGluLysValLysGlyAspAspHisAsp 279  
 QY |||||  
 712 GACACCTCCGATAGTACACGATGCTGCTGACCTTGAAAGAGCTTCAGAACTCCCTTCG 771  
 DDb |||||  
 280 AspThrSerAspSerAspSerAspGlyLeuThrLeuLysGluLeuGlnAsnArgLeuArg 299  
 QY |||||  
 772 AGAAGCGGAACAGGAGCCCACTGAGAGCCCTCTGAAGGGATCCAGAGTCCCTCGCG 831  
 DDb |||||  
 300 ArgLysArgGluGlnLuproThrGluArgProLeuLysGlyIleGlnSerArgLeuArg 319  
 QY |||||  
 832 AAGAGCGCGGAGGAGGTCCCGCGACACTGTGGCTCCGAGGCGCAGTGACACTGTG 891  
 DDb |||||  
 320 LysLysArgArgGluGluGlyProAlaGluThrValGlySerGluAlaSerAspThrVal 339  
 QY |||||  
 892 GAGGCGCTCTGCCCAGTAAGCAGGAGCCCGAGAACGATCAGGGGTTGTGCCAGGCT 951  
 DDb |||||  
 340 GluGlyValLeuProSerLysGlnGluProGluAsnAspGlnGlyValValSerGlnAla 359  
 QY |||||  
 952 GGGAAAGATCAGAGAGTAACTGGAGGGAAGCGCTCAGAGCATCAAGATGAG 1011  
 DDb |||||  
 360 GlyLysAspArgGluSerLysLeuGluGlyLysAlaAlaGlnAspIleLysAspGlu 379  
 QY |||||  
 1012 GAGCTGAGACTTGGCCGACGAGCCTGAATGTGAGGGTTACGACCCCAAGCCCTG 1071  
 DDb |||||  
 380 GluProGlyAspLeuGlyArgProLysProGluCysGluGlyTyraAspProAsnAlaLeu 399

QY 1072 TATTGCATTTCCGCCAGCCCTCAACAACAGAGTTTATGATTTTCTGTGACCGGTGTGAA 1131  
 DDb |||||  
 400 TyrCysIleCysArgGlnProIleAsnAsnArgPheMetIleCysCysAspArgCysGlu 419  
 QY |||||  
 1132 GAATGGTTTCATGGCGATTGTGTGGCATTTCTGAGGCTCGAGGAGGCTTTTGGAAAGG 1191  
 DDb |||||  
 420 GluTrpPheHisGlyAspCysValGlyIleSerGluAlaArgGlyArgLeuLeuGluArg 439

QY 1192 AATGGGGAAGACTATATCTGCCCAAACTGCACCACTTCTGCAAGTGCAGGATGAGACTCAT 1251  
 DDb |||||  
 440 AsnGlyGluAspTyrIleCysProAsnCysThrIleLeuGlnValGlnAspGluThrHis 459  
 QY |||||  
 1252 TCAGAAAGCGCAGATTCAGCAGGAAGCTTAATGGAGACCT---GGAGATGCTGATGGCACC 1308  
 DDb |||||  
 460 SerGluThrAlaAspGlnGlnGluAla\*\*\*MetGlyAspLeuGlyAspAlaAspGlyThr 479

QY 1309 GATTGTACAAGTATAGGAACAATAGACGAGAAG---TCTAGCGAAGACCAAGGGATA 1362  
 DDb |||||  
 480 AspPheThrSerIleGlyAsnIleArgSerArgSerLeuSerArgArgProLysGlyLeu 499  
 QY |||||  
 1363 AAGGTAGAAATTGAGAAAGCTGCAAAATCCAAAGTGGCAAGAACTCAAGATCTTCCAG 1422  
 DDb |||||  
 500 LysGlyArgIleGluLysAlaAlaAsnProSerGlyLysLysGluLeuLysIlePheGln 519

QY 1423 CCT----- 1425  
 DDb |||||

QY 520 ProValLeuGluIleAlaValSerArgSerIleSerAlaPheThrLeuLeuHisCysIle 539  
 DDb |||||

QY 1426 -----GTGATAGAGCGCCTGGTCCCTCAAAATGATATGCGCCCGGTGTGTAC 1476  
 DDb |||||

QY 540 SerCysLysValIleGluAlaProGlyAlaSerLysCysIleGlyProGlyCysCysHis 559  
 QY |||||

QY 1477 GTGGCGCAGCCGACTCGGTGCTACTGCGAGTAATGACTTATCTCAACACAGCCCGCAGG 1536  
 DDb |||||

QY 560 ValAlaGlnProAspSerValTyrCysSerAsnAspCysIleLeuLysHisAlaAlaAla 579  
 QY |||||

QY 1537 ACAATGAAGTTTCTAAGCTCAGGTAAGAACAGAACGCAAGCCCTAAAGAAAGATGAAG 1596  
 DDb |||||

QY 580 ThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysProLysGluLysMetLys 599  
 QY |||||

QY 1597 ATGAAGCCAGAGAGCCCGAGTCTCCGAAATCGCGTGTCTCAG----- 1638  
 DDb |||||

QY 600 MetLysProGluLysProSerLeuProLysCysGlyAlaGlnProValGlnTrpLeuLeu 619  
 QY |||||

QY 1639 -----GCAGGTATTAATAATCTCTCTGTGTGCACAG 1668  
 DDb |||||

QY 620 LeuLeuSerSerTrpGlnLeuGlyThrSerSerGlyLeuGlnVal---ThrLeuCysGln 638  
 QY |||||

QY 1669 AGACCAAGTCCAGAAAAAAGAGACCCAGTGAAGAGGAGTGTGTGTCTCCCTCGCGG 1728  
 DDb |||||

QY 639 ProProPheThrGluHisLys-----TrpLeuLysPheSerValLeu-----His 653  
 QY |||||

QY 1729 AGTGAAGCACTCGGGAAGGAAGCAGCT-----TGTGAGAGCAGCAGCCGCTCGTGGCG 1782  
 DDb |||||

QY 654 GlyGluThrAlaGlyProProArgAlaHisArgIleArgLysLeuAsnLysTyrTrpGly 673  
 QY |||||

QY 1783 AGCGATCAATATCAATGCAATGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1842  
 DDb |||||

QY 674 AlaCysHisHisCysMetProLeuArgProLeu----- 684  
 QY |||||

QY 1843 TTGTATAATGATGATCATCACTAGGGTGGCCCTCGACCCCTCCCGTTCTTTCTGCG 1902  
 DDb |||||

QY 685 -----ProCysLeuProGlyTrp 690  
 QY |||||

QY 1903 ATAGCCATCCCTGGGCTGTCCAGGACTG---GGAGTTGACAGCTTTGTGTAAAGTGTAT 1959  
 DDb |||||

QY 691 -----ProArgSerSerProSerLeuValGlyAlaGlySerLeu----- 703  
 QY |||||

QY 1960 CACAGACACCGCTGCAACCATCAGCGGAGGAGCAGAGCCCATGTCCAGGATGCTCTGCT 2019  
 DDb |||||

QY 704 ProArgHisAlaValHis-----LeuAsnGluProCysProLeu 716  
 DDb |||||

QY 2020 GCCCTGTGTCCA 2031

DB	717 AlaileCysPro 720	:
RESULT 5		
AAM79235		
ID	AAM79235 standard; protein; 647 AA.	
XX	AC	
XX	AC	
XX	XX	
DT	06-NOV-2001 (first entry)	
DE	Human protein SEQ ID NO 1897.	
XX		
KW	Human; cytokine; cell proliferation; cell differentiation; gene therapy; vaccine; peptide therapy; stem cell growth factor; haematopoiesis; tissue growth factor; immunomodulatory; cancer; leukaemia; nervous system disorder; arthritis; inflammation.	
KW		
XX		
OS	Homo sapiens.	
XX		
PN	WO200157190-A2.	
XX		
PD	09-AUG-2001.	
XX		
PF	05-FEB-2001; 2001WO-US004098.	
XX		
PR	03-FEB-2000; 2000US-00496914.	
PR	27-APR-2000; 2000US-00560875.	
PR	20-JUN-2000; 2000US-00598075.	
PR	19-JUL-2000; 2000US-00620325.	
PR	01-SEP-2000; 2000US-00654936.	
PR	15-SEP-2000; 2000US-00663561.	
PR	20-OCT-2000; 2000US-00693325.	
PR	30-NOV-2000; 2000US-00728422.	
XX		
PA	(HYSE-) HYSEQ INC.	
XX		
PI	Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;	
PI	Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;	
PI	Xue AJ, Yang Y, Wejhrman T, Goodrich R;	
XX		
DR	WPI; 2001-476203/51.	
DR	N-PSDB; AAK52368.	
XX		
PT	Nucleic acids encoding polypeptides with cytokine-like activities, useful in diagnosis and gene therapy.	
PT		
XX		
PS	Claim 20; Page 4293-4294; 6221pp; English.	
XX		
CC	The invention relates to polynucleotides (AAK51456-AAK53435) and the encoded polypeptides (AAM78323-AAM80302) that exhibit activity relating to cytokine, cell proliferation or cell differentiation or which may induce production of other cytokines in other cell populations. The polynucleotides and polypeptides are useful in gene therapy, vaccines or peptide therapy. The polypeptides have various cytokine-like activities, e.g. stem cell growth factor activity, haematopoiesis regulating activity, tissue growth factor activity, immunomodulatory activity and activin/inhibin activity and may be useful in the diagnosis and/or treatment of cancer, leukaemia, nervous system disorders, arthritis and inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the sequence listing were missing at the time of publication	
XX		
SQ	Sequence 647 AA;	
Alignment Scores:		
Pred. No.:	2,376-171	Length: 647
Score:	2311.00	Matches: 441
Percent Similarity:	92.45%	Conservative: 0
Best Local Similarity:	92.45%	Mismatches: 0
Query Match:	49.29%	Indels: 36
DB:	4	Gaps: 1

QY 1501 TGCAGTAATGACTGATCTCTCAACACGCGCGAGGACCAATGAAGTTTCTTAAGCTCAGGT 1560  
 Db |||||  
 QY 361 CysSerAspSpyslleuylsHlslalalalathrMetLysPheLeuSerSerGly 380  
 Db |||||  
 QY 1561 AAAGAACAAGACCAAGCCTAAAGAAAGATGAAGATGAAGCCAGAGCCAGCTCTT 1620  
 Db |||||  
 QY 381 LysGluGlnLysProLysProLysGluLysMetLysMetLysProGluLysProSerLeu 400  
 Db |||||  
 QY 1621 CCGAATGCGGTCTCAGCAGCTATTAATACTCTTCTGTGCACAAAGAGACCACTCCA 1680  
 Db |||||  
 QY 401 ProLysCysGlyAlaGlnAlaGlylleLyslleSerSerValHlsLysArgProAlaPro 420  
 Db |||||  
 QY 1681 GAAAAAAGAGACCACTGAGAACAGGAGGTGGTCTCTCCGCGGAGTCAAGCACTC 1740  
 Db |||||  
 QY 421 GluLysLysGluThrThrValLysLysAlaValValProAlaArgSerGluAlaLeu 440  
 Db |||||  
 QY 1741 GGAAGAGAGACCTGTGTGAGACGACGCGCTGTGTGGCGGAGCATCACATTAAT 1800  
 Db |||||  
 QY 441 GlyLysGluAlaAlaCysGluSerSerThrProSerTrpAlaSerAspHisAsnTrpAsn 460  
 Db |||||  
 QY 1801 GCAGTAAAGCCAGAAAGAGCTGTCTCTCCCTCGCGCTCACTGTGTATATAA 1851  
 Db |||||  
 QY 461 AlaValLysProGluLysThrAlaAlaProSerProSerLeuLeuTrpLys 477  
 Db |||||  
 RESULT 6  
 ID AAY67580  
 AC AAY67580 standard; protein; 614 AA.  
 XX AAY67580;  
 DT 19-JUN-2000 (first entry)  
 XX Murine death inducer-obliterator 1 (DIO-1) polypeptide.  
 DE Death inducer-obliterator 1; DIO-1; cell death; cancer; tumour; mouse;  
 KW autoimmune disease; cytostatic; immunosuppressive; antidiabetic;  
 KW antirheumatic; antiinflammatory; antiproliferative.  
 XX Mus sp.  
 XX Key  
 FH Region  
 FT Location/Qualifiers  
 FT 162..170  
 FT /note= "NLS sequence"  
 FT 182..190  
 FT /note= "NLS sequence"  
 FT 262..285  
 FT /note= "zinc finger motif"  
 FT 290..315  
 FT /note= "zinc finger motif"  
 XX WO200015787-A1.  
 XX 23-MAR-2000.  
 XX 10-SEP-1999; 99WO-GB003019.  
 XX 10-SEP-1998; 98SE-00003069.  
 XX 17-SEP-1998; 98US-0100873P.  
 XX (CNSJ ) CONSEJO SUPERIOR INVESTIGACIONES CIENTIF.  
 XX (BANN/) BANNERMAN D G.  
 XX Alonso CM, Domingo DG, Grandien A, Leonardo E, Martinez P;  
 XX WPI; 2000-271426/23.  
 XX N-PSDB; AAZ90579.  
 XX New DNA encoding human and murine death inducer-obliterator 1  
 XX polypeptides, useful in the treatment of cancer, autoimmune diseases,  
 XX diabetes, rheumatoid arthritis, benign tumors, malignant tumors and  
 XX hyperproliferative skin disorders.

PS Claim 8; Fig 1D; 27pp; English.  
 XX The invention provides nucleic acids encoding the human and murine death  
 CC inducer-obliterator 1 (DIO-1) polypeptides. The polypeptides can be  
 CC expressed by standard recombinant methodology. The DIO-1 polypeptides,  
 CC agonists and antagonists are used as a medicament for treating diseases,  
 CC characterized by an alteration in cell death or by hyperproliferation,  
 CC e.g. cancer, autoimmune diseases, diabetes, rheumatoid arthritis, benign  
 CC tumours, malignant tumours or hyperproliferative skin disorders. They are  
 CC also useful in the treatment of metabolic, proliferative or inflammatory  
 CC conditions. The present sequence represents the murine DIO-1 polypeptide  
 XX Sequence 614 AA;  
 SQ  
 Alignment Scores:  
 Pred. No.: 3,79e-159 Length: 614  
 Score: 2154.50 Matches: 449  
 Percent Similarity: 76.96% Conservative: 42  
 Best Local Similarity: 70.38% Mismatches: 112  
 Query Match: 45.95% Indels: 36  
 Gaps: 12  
 US-09-787-016A-1 (1-2610) x AAY67580 (1-614)  
 QY 265 AFGACACGACAAAGCGACCCCGAGCAATGAGGAGGACCTAAGGCCATCAAAACCCACGAC 324  
 Db |||||  
 QY 1 MetAspAspLysGlyHisLeuSerAsnGluGluAlaProLysAlaIleLysProThrSer 20  
 QY 325 AAAGAGTTCCAGAAACATGGGTTTCGAAGGACCACTATCGCCAGCGAGGAGCGCGCA 384  
 Db |||||  
 QY 21 LysGluPheArgLysThrTrpGlyPheArgArgThrIleAlaLysArgGluGlyAla 40  
 QY 385 GGGAGCGGCGGAGCTGACCCACTGAGCGCCGCCACCCACACGACGACGAGCTGGCGCTGTCC 444  
 Db |||||  
 QY 41 GlyAspThrGluAlaAspProSerGluGlnGlnPro-----GlnGlnHisAsnLeuSer 58  
 QY 445 CTGCGCGCAGTGGGAGGAGCGACCCAAAGCCACTGAGCGGTGGAGCAGCTTCCTGACCATT 504  
 Db |||||  
 QY 59 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGluPheLeuThrThr 78  
 QY 505 GCGCGCGCGCGCGAGGAGGAGCATGCTGTCTCTCGAGGATTTCTGTGAGCGCCACG 564  
 Db |||||  
 QY 79 ValArgArgArgGlyLysLysAsnValProValSerLeuGluAspSerSerGluProThr 98  
 QY 565 TCTGTCCCGCCGACAGACCGCGAGCAGCTCCGAGGCGGAGGTGGAAAGCGCTTCTGAG 624  
 Db |||||  
 QY 99 SerSerThrValThrAspValGluThrAlaSerGluGlySerValGluSerSerSerGlu 118  
 QY 625 ACCAGAGCGCGCCCGCAGTCTGTTCCACAGCTGTGAAGAAACGACGACCTCTTCTGAA 684  
 Db |||||  
 QY 119 IleArgSerGlyProValSerAspSerLeuGly---LysGluHisProLysSerSerGlu 137  
 QY 685 AAGGTGAAGAGGAGGATCAACCAATGACACTCTCGATGATGACAGCGATGCGCTGACC 744  
 Db |||||  
 QY 138 LysAlaLysGlyGlyGluGluGluAspThrSerAspSerAspSerAspGlyLeuThr 157  
 QY 745 TTGAAGAGCTTCAGATCGCTTCGACGAGAGCGGAAACAGGAGCGCCACTGAGAGCGCC 804  
 Db |||||  
 QY 158 LeuLysGluLeuGlnAsnArgLeuArgLysArgGluGluGluGluProValGluArgSer 177  
 QY 805 CTGAAGAGGATCCAGAGTGCCTCGCGGAAGAACCGCGGAGGAGGCTCCCGCGGAGCT 864  
 Db |||||  
 QY 178 LeuArgGlySerGlnAsnArgLeuArgLysLysArgGluGluGluAspSerAlaGluThr 197  
 QY 865 GTGGGCTCCGAGCGCCAGTGCATCTGTGGGGCGCTCTGCCC---AGTAGAGCAGGAGCC 921  
 Db |||||  
 QY 198 ---GlySerValGlnIleGlySerAlaGluGlnAspArgProLeuCysLysGlnGluPro 216  
 QY 922 GAGAACGATCAGGGGTTGTCTCCAGGCTGGGAAAGATGACAGAGAGTAAATGTTGAG 981  
 Db |||||  
 QY 217 GluAlaSerGlnGlyProValSerGlnSerGluThrAspAspIleGluAsnGlnLeuGlu 236  
 QY 982 GGAAAGCGCGCTCAGGACATCAAGATGAGAGCTGGAGACTTGGGCCCGACCGAGCCT 1041

```

237 GlyLysAlaThrGlnGlyAsnThrGluGluAsnProArgGluAlaGlyLysProLysPro 256
1042 GAATGTGAGGTTTACGACCCCAACCGCCCTGTATTGCAATTCGCCGAGCTCACAAAC 1101
257 GluCysGluValTyrAspProAsnAlaLeuTyrCysIleCysArgGlnProHisAsnAsn 276
1102 AGCTTTATGATTTGCTGTCGACCCCTGGAAGATGTTTCATGCGCATTTGTGGCATT 1161
277 ArgPheMetIleCysCysAspAryCysGluGluThrPheHisGlyAspCysValGlyIle 296
1162 TCTGAGCTCGAGGAGCGCTTTTGGAAAGAAATGGGGAAGACTATATCTGCCCAACTGC 1221
297 SerGluAlaArgGlyArgGluLeuGluArgAsnGlyGluAspTyrIleCysProAsnCys 316
1222 ACATTTCTGAGTGCAGATGAGATCATTCAAAACGCCACATCAGCAGGAGACTAA 1281
317 ThrIleLeuGlnValGlnAspGluThrAsnGlySerAlaThrAsnGluGlnAspSerGly 336
1282 TGGAGCTCGAGATGCTGATGCGACCGATTGTACAGTATAGAAACAATAGACAGAGAAG 1341
337 CysArgSerValGlyAlaAspGlyThrAspCysThrSerIleGlyThrValGluGlnLys 356
1342 TCTAGCGAAGACACAGGATTAAGGTAGATTAATTCAGAAAGCTGCAATCCAAATGCGCAG 1401
357 SerGlyGluAspGlnGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLys 376
1402 AAGAACTCAAGATCTTCAGCCTGTGATAGAGCGCTGCTCCTCAAAATGATTGGC 1461
377 LysLysLeuLysIlePheGlnProValValGluAlaProGlyAlaProLysCysIleGly 396
1462 CCGCGTGTCTGTCAGTGGCGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1521
397 ProGlyCysSerSerValAlaGlnProAspSerValTyrCysSerAsnAspCysIleLeu 416
1522 AATACCCCGACGACGATGCTTCAAGCTCAGTAAAGACAGACAGACAGACAGCCT 1581
417 LysHisAlaAlaAlaThrMetArgPheLeuSerSerGlyLysGluGlnLysThrLysPro 436
1582 AAGAAAGATGAAGATGAAGACAGACAGACAGACAGCCTGCTCGAAATGCGTCTCAGGCA 1641
437 LysGluLysValLysThrLysProGluLysPheSerLeuProLysCysSerValGlnVal 456
1642 GGTATTAATCTTCTGTCACAGACAGACAGACAGCCTCCAGAAAAGACAGACAGCAGTG 1701
457 GlyIleLysIleSerSerValHisLysArgLeuAlaSerGluLysArgGluAsnProVal 476
1702 AAGAGGAGTGGTGGTCCCTGCGCAGTGAAGCACTCGCGAGGAGGAGGAGGAGTGTGAG 1761
477 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 495
1762 AGCAGCAGCGCTGCTGGGAGCGATCACAATTAATGATGATGATGATGATGATGATGATG 1821
496 SerSerThrProSerTrpAlaSerAspHisAsnTyrAsnAlaValLysProGluLysPro 515
1822 GCTGCTCC-----TGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1872
516 GluLysProThrAlaLeuSerProThrLeuLeuSerLysCysThrThrHisProLysAla 535
1873 GGCCT-CTTGACCCCTCCCGTCTTCTTCTGATAGCCATCC-----SerHisLeuGlyGlyCysLeuGly 549
536 GlyPheProGlyPro-----SerHisLeuGlyGlyCysLeuGly 549
1920 CTGTCCAGGACT-----GGAGTTCAGCTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1970
550 LeuSerArgThrArgValLeuValLeuValLeuValLeuValLeuValLeuValLeuVal 569
1971 GCTGCACCATCAGCGGAGCAGACAGCAGCAGTCTCAGGATGCTCTGCTGCTGCTGCTGCTC 2030
570 Ala-----ArgSerArgTyrGlnAspAlaSerGlyProGlnValPhe 583
2031 ATCCCTAGTCTGCAGGACTTCTGTCACCTGTTTCCAAAGCTGTAAACCTCACTGCTGA 2090
:::|||||
```

```

584 LeuProSerLeuTrpSerLeuSerGlyTrpPheLeuLysSerCys----- 598
2091 AGCTTCACTTAATGATTCATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTT 2144
599 ---ValGlyLeuMetLeuAla---IleSerTyrPheSerPheArgProIleP 614

RESULT 7
ABG22388
ID ABG22388 standard; protein; 330 AA.
XX
XX ABG22388;
XX
XX 18-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #22379.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US008631.
XX
XX 31-MAR-2000; 2000US-00540217.
XX 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX N-PSDB; AAS86575.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX
XX Claim 20; SEQ ID NO 52747; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostics as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantitating a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful in medical imaging
XX of sites expressing (II). (I) and (II) are useful for treating disorders
XX involving aberrant protein expression or biological activity. The
XX polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
XX amino acid sequences of the invention. Note: The sequence data for this
XX patent did not appear in the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 330 AA;
XX
XX Alignment Scores:
XX Pred. No.: 3,74e-112 Length: 330
XX Score: 1551.00 Matches: 290
XX Percent Similarity: 88.96% Conservative: 0
XX Best Local Similarity: 88.96% Mismatches: 0
```

Query Match: 33.08%		Indels: 36
DB: 4		Gaps: 1
US-09-787-016A-1 (1-2610) x ABG22388 (1-330)		
Qy	967	GAGAGTAAGTTGAGGGAAGGCGGTCTAGGACATCAAGATGAGAGCTGAGACTTG 1026
Ds	1	GluserLysLeuGluGlyLysAlaAlaGlnAspIleLysAspGluGluProGlyAspLeu 20
Qy	1027	GGCGGACGAGCTGAATGAGGTAGACCCAGCCCTGATTGCTATTCCTCCGC 1086
Ds	21	GlyArgProLysProGluCysGluGlyLysAspProAsnAlaLeuLysCysIleCysArg 40
Qy	1087	CAGCCTCACACAAACAGTTATGATTGCTGACCGCTGTGAAGATGTTTCATGCGC 1146
Ds	41	GlnProHisAsnAsnArgPheMetIleCysCysAspArgCysGluGluLysPheHisGly 60
Qy	1147	GATTGCTGGCATTCTGAGGCTCGAGGAGCTTTTGCAGAGGATGGGAGACTAT 1206
Ds	61	AspCysValGlyIleSerGluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyr 80
Qy	1207	ATCTGCCCAACTGCACATTCTGCAAGTGCAGATGAGACTCATTCAGAAACGGCAGAT 1266
Ds	81	IleCysProAsnCysThrIleLeuGlnValGlnAspGluThrHisSerGluThrAlaAsp 100
Qy	1267	CAGCAGAACTAAATGAGACTGGAGATGCTGATGCGACCGATGTTACAGTATAGGA 1326
Ds	101	GlnGlnGluAlaLysTyrArgProGlyAspAlaAspGlyThrAspCysThrSerIleGly 120
Qy	1327	ACAATGAGCAGAGTCTAGCAGACCAAGGATTAAGGCTAGAAATTGAGAAAGCTGCA 1386
Ds	121	ThrIleGluGlnLysSerSerGluAspGlnGlyIleLysGlyArgIleGluLysAlaAla 140
Qy	1387	AATCAAGTGCAGAGAACTCAAGATCTTCAGCCT----- 1425
Ds	141	AsnProSerGlyLysLysLysLeuLysIlePheGlnProGlyProGlyProValProThr 160
Qy	1425	----- 1425
Ds	161	GlnLeuProValLeuTyrGlnValLeuGluIleAlaValSerArgSerIleSerAlaPhe 180
Qy	1426	-----GTGATGAGCGCGCTGGTGCCTCAAAATGATT 1458
Ds	181	ThrLeuLeuHisCysIleSerCysLysValIleGluAlaProGlyAlaSerLysCysIle 200
Qy	1459	GGCCCCGGTCTGCTACGTCGGCGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1518
Ds	201	GlyProGlyCysCysHisValAlaGlnProAspSerValTyrCysSerAsnAspCysIle 220
Qy	1519	CTCAACACCGCGCAGCAGCAATGAAGTTCTTAAGCTCAGGTAAAGAACAGAGCCAAAG 1578
Ds	221	LeuLysHisAlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLys 240
Qy	1579	CCTAAGAAAGATGAGATGAGCCAGAGCCCGCTCTCCGAATCGGTCCTCAG 1638
Ds	241	ProLysGluLysMetLysMetLysProGluLysProSerLeuProLysCysGlyAlaGln 260
Qy	1639	GCAGTATTAATCTCTCTGTCACAGAGACAGCTCCAGAAAAAAGAGACACACA 1698
Ds	261	AlaGlyIleLysIleSerSerValHisLysArgProAlaProGluLysLysGluThrThr 280
Qy	1699	GTGAAGAGGAGGTGGTGGTCCCTCGCGGAGTGAAGCATTGCGGAGAGGAGAGCTTGT 1758
Ds	281	ValLysLysAlaValValProAlaArgSerGluAlaLeuGlyLysGluAlaAlaCys 300
Qy	1759	GAGAGCAGCGCTGCTGGCGGAGCGATCACATTAATGAGTAAAGCCAGAAAG 1818
Ds	301	GluserSerThrProSerTrpAlaSerAspHisAsnTyrAsnAlaValLysProGluLys 320
Qy	1819	ACTGCTGCTCCCTCGCGC 1836
Ds	321	ThrAlaAlaProSerPro 326

RESULT 8	
AAB43724	
ID	AAB43724 standard, protein; 181 AA.
XX	
AC	AAB43724;
XX	
DT	08-FEB-2001 (first entry)
XX	
DE	Human cancer associated protein sequence SEQ ID NO:1169.
XX	
KW	Human; cancer associated gene; cancer antigen; detection; cancer;
KW	diagnosis; cytostatic; proliferative; vulnarary; immunomodulator;
KW	antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;
KW	antiinflammatory; antithyroid; antiallergic; antibacterial; cardiac;
KW	dermatological; neuroprotective; thrombolytic; coagulant; neoplastic;
KW	vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
KW	immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW	allergic reaction; graft versus host disease; organ rejection;
KW	haemostatic; thrombolytic; cardiovascular disorder; infection;
KW	neurological disease; drug screening.
XX	
OS	Homo sapiens.
XX	
PN	WO200005350-A1.
XX	
PD	21-SEP-2000.
XX	
PF	08-MAR-2000; 2000WO-US005882.
XX	
PR	12-MAR-1999; 99US-0124270P.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	
PI	Rosen CA, Ruben SM;
XX	
DR	WPI; 2000-587533/55.
DR	N-PSDB; AAC77933.
XX	
PT	Novel isolated nucleic acids comprising sequences encoding peptides
PT	useful for treating or diagnosing e.g. cancer.
XX	
PS	Claim 11; Page 1790-1791; 2352pp; English.
XX	
CC	AAC77607 to AAC78448 encode the human cancer associated proteins given in
CC	AAB43398 to AAB44239. The proteins can have activities based on the
CC	tissues and cells the genes are expressed in. Example of activities
CC	include: cytostatic; proliferative; vulnarary; immunomodulator;
CC	antidiabetic; antiasthmatic; antirheumatic; antiarthritic;
CC	antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
CC	dermatological; neuroprotective; cardiac; thrombolytic; coagulant;
CC	neoplastic; vasotropic; antipsoriatic and antiangiogenic. The
CC	polynucleotides and polypeptides can be used for preventing, treating or
CC	ameliorating medical conditions and diagnosing pathological conditions.
CC	Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC	the present invention may be used to treat immune disorders by activating
CC	or inhibiting the proliferation, differentiation or mobilisation of
CC	immune cells, to treat disorders of haematopoietic cells, autoimmune
CC	disorders, allergic reactions, graft versus host disease, and organ
CC	rejection, modulate haemostatic or thrombolytic activity, modulate
CC	inflammation, cancers, cardiovascular disorders, neurological disease and
CC	bacterial or viral infections. The peptides, nucleotides, antibodies,
CC	agonists and antagonists may be also be used in drug screens. AAC78449 to
CC	AAC78457 and AAB44240 represent sequences used in the exemplification of
CC	the present invention
XX	
SQ	Sequence 181 AA;
Alignment Scores:	
Pred. No.:	5,63e-66 Length: 181
Score:	98.00 Matches: 179
Percent Similarity:	98.35% Conservatives: 0
Best Local Similarity:	98.35% Mismatches: 2
Query Match:	20.43% Indels: 2

```

DB: 3 Gaps: 0
US-09-787-016A-1 (1-2610) x ABA43724 (1-181)
QY 998 ACATCAAGATGAGAGCTTGGAGACTTG-GCCCGACCGAACCCCTGAATGTGAGGTGAC 1056
Db 1 ThrSerLysMetArgSerLeuGluThrLeuGlyArgProLysProGluCysGluGlyTyr 20
QY 1057 GACCCCAAGCCCTGATTTGCTTGGCGCGAGCTCACAACAACAGGTTTATGATTGTC 1116
Db 21 AspProAsnAlaLeuTyrCysIleCysArgGlnProHisAsnAsnArgPheMetIleCys 40
QY 1117 TGTGACCGCTGTCAAGAAATGTTTCATGCGCGATTGTGTGGCAATTTCTGAGCTCGAGG 1176
Db 41 CysAspArgCysGluGluTrpPheHisGlyAspCysValGlyIleSerGluAlaArgGly 60
QY 1177 AGCTTTTGGAAAGGATGGGAGACTATATCTGCCAAACTGCACCATCTGCAAGTG 1236
Db 61 ArgLeuLeuGluArgAsnGlyGluAspTyrIleCysProAsnCysThrIleLeuGlnVal 80
QY 1237 CAGGATGAGACTCATTCAAAAACGGCAGATCAGCAGAGCTAAATGGAGACTTGAGAT 1296
Db 81 GlnAspGluThrHisSerGluThrAlaAspGlnGlnAlaIleTyrPargProGlyAsp 100
QY 1297 GCTGATGGACCGCATTTGACAGTATAGGAACAATAGACAGAGTCTAGCGAGACCAA 1356
Db 101 AlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSerSerGluAspGln 120
QY 1357 GGGTAAGGGTAGATTGAGAAAGCTGCAAAATCCAAATGGCAAGAGAACTCAAGATC 1416
Db 121 GlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLysLysLysLysIle 140
QY 1417 TTCCAGCTGTGATAGAGCGCTGTGTGCTCAAAATGTATTGGCCCGGCTGTCTAC 1476
Db 141 PheGlnProValIleGluAlaProGlyAlaSerLysCysIleGlyProGlyCysHis 160
QY 1477 GTGGCCACCGCAGCTCGGTGTACTGTCAGTAAATGACTGTATCTCAACACGCCGAGG 1536
Db 161 ValAlaHis-ProThrArgCysThrAlaValMetThrValSerSerAsnThrProGlnAr 180
QY 1537 ACAA 1540
Db 180 gGln 181
RESULT 9
ID ABG22387
AC ABG22387 standard; protein; 167 AA.
XX ABG22387;
XX 18-FEB-2002 (first entry)
XX Novel human diagnostic protein #22378.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US008631.
XX 31-MAR-2000; 2000US-00540217.
XX 23-AUG-2000; 2000US-00649167.
XX (HYSB-) HYSBQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.

```

```

DR N-PSDB; AAS86574.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
PS Claim 20; SEQ ID NO 52746; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
sequences. (I) is useful as hybridisation probes, polymerase chain
reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
and in recombinant production of (II). The polynucleotides are also used
in diagnostics as expressed sequence tags for identifying expressed
genes. (I) is useful in gene therapy techniques to restore normal
activity of (II) or to treat disease states involving (II). (II) is
useful for generating antibodies against it, detecting or quantitating a
polypeptide in tissue, as molecular weight markers and as a food
supplement. (III) and its binding partners are useful in medical disorders
of sites expressing (II). (I) and (II) are useful for treating disorders
involving aberrant protein expression or biological activity. The
polypeptide and polynucleotide sequences have applications in
diagnostics, forensics, gene mapping, identification of mutations
and to produce other types of data and products dependent on DNA and
amino acid sequences. ABG0010-ABG3037 represent novel human diagnostic
amino acid sequences of the invention. Note: The sequence data for this
patent did not appear in the printed specification, but was obtained in
electronic format directly from WIPO at
ftp.wipo.int/pub/published_pct_sequences
XX Sequence 167 AA;
SQ
Alignment Scores:
Pred. No.: 2,636-61 Length: 167
Score: 898.00 Matches: 167
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 19.15% Indels: 0
DB: 4 Gaps: 0
US-09-787-016A-1 (1-2610) x ABG22387 (1-167)
QY 1174 GGGAGCTTTTGGAAAGGAATGGGGAAGACTATATCTGCCAAACTGCACCATTCTGCAA 1233
Db 1 GlyArgLeuLeuGluArgAsnGlyGluAspTyrIleCysProAsnCysThrIleLeuGln 20
QY 1234 GTCCAGGATGAGACTCATTTCAGAAACGGCAGATCAGCAGAGCTAAATGGAGACTGGA 1293
Db 21 ValGlnAspGluThrHisSerGluThrAlaAspGlnGlnAlaLysTrpArgProGly 40
QY 1294 GATGCTGATGGCAGCGATTGTACAGTATAGGAACAATAGACAGCAAGTCTAGCGAGAC 1353
Db 41 AspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSerSerGluAsp 60
QY 1354 CAAAGGATAAAGGTAGATTGAGAAAGCTGCAAAATCCAAATGGCAAGAGAACTCAAG 1413
Db 61 GlnGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLysLysLysLeuLys 80
QY 1414 ATCTTCCAGCTGTGATAGAGCGCTGTGTGCTCAAAATGTATTGGCCCGGCTGTCT 1473
Db 81 IlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGlyProGlyCysCys 100
QY 1474 CACGTGGCGCAGCCGACTCGGTGTACTGTCAGTAAATGACTGTATCTCAAAACGCCGCA 1533
Db 101 HisValAlaGlnProAspSerValTyrCysSerAsnAspCysIleLeuLysHisAlaAla 120
QY 1534 GCGACAATGAAGTTTCTAAGCTCAGGTAAAGAACAGAGCCAAAGCTTAAGAAAGATG 1593
Db 121 AlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysProLysGluLysMet 140
QY 1594 AAGATGACAGCCAGAGAGCCAGTCTTCGGAATCGGTGCTCAGCAGGTATTAAATC 1653

```



Db	141	LysMetLysProGluLysProSerLeuProLysCysGlyAlaGlnAlaGlyIleLysIle	167
Qy	1654	TCTTCTGTGCACACAGACCA	1674
Db	161	SerSerValHisLysArgPro	167
RESULT 10			
ABP05496			
ID	ABP05496	standard; protein; 108 AA.	
XX			
AC	ABP05496;		
XX			
DT	24-JUN-2002	(first entry)	
XX			
DE	Human ORFX protein sequence SEQ ID NO:10974.		
XX			
KW	Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;		
XX	hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;		
KW	degenerative disorder; osteoarthritis; neurodegenerative disorder;		
XX	cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;		
KW	hypertension; hypothyroidism; cholesterol ester storage disease;		
XX	immune deficiency; immune disorder; infectious disease;		
KW	autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;		
XX	myasthenia gravis.		
OS	Homo sapiens.		
XX			
PN	WO2001092523-A2.		
XX			
PD	06-DEC-2001.		
XX			
PF	29-MAY-2001; 2001WO-US010836.		
XX			
PR	30-MAY-2000; 2000US-0206132P.		
XX			
PR	29-AUG-2000; 2000US-0228718P.		
XX			
PA	(CURA-) CURAGEN CORP.		
XX			
PI	Shimkets RA, Leach MD;		
XX			
DR	WPI; 2002-106308/14.		
XX			
DR	N-PSDB; ABN21248.		
XX			
PT	Novel human polypeptides and polynucleotides useful for diagnosing,		
XX	preventing and treating cardiovascular disease, neurodegenerative,		
PT	hyperproliferative disorders and autoimmune disorders.		
XX			
PS	Disclosure; SEQ ID NO 10974; 1037pp; English.		
XX			
CC	The present invention describes substantially purified human proteins		
XX	(referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1		
CC	in the specification). ABN15762 to ABN27252 encode the human ORFX		
XX	proteins given in ABP00010 to ABP11500. ORFX proteins are useful for		
CC	treating or preventing a pathology associated with an ORFX-associated		
XX	disorder in humans, and in the manufacture of a medicament for treating a		
CC	syndrome associated with ORFX-associated disorder. ORFX polynucleotide		
XX	sequences can be used in gene therapy. ORFX sequences can be used in the		
CC	treatment of cancer, hyperproliferative disorders, cirrhosis of liver,		
XX	psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,		
CC	osteoarthritis, neurodegenerative disorders, disorders related to organ		
XX	transplantation, cardiovascular diseases, diabetes mellitus, systemic		
CC	lupus erythematosus, hypertension, hypothyroidism, cholesterol ester		
XX	storage disease, various immune deficiencies and disorders, infectious		
CC	diseases, autoimmune disorders such as multiple sclerosis, rheumatoid		
XX	arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host		
CC	disease and autoimmune inflammatory eye disease. ORFX proteins are also		
XX	useful for treating burns, incisions, ulcers, for treating osteoporosis,		
CC	bone degenerative disorders, or periodontal disease, and for gut		
XX	protection or regeneration and treatment of lung or liver fibrosis,		
CC	reperfusion injury in various tissues and conditions resulting from		
XX	systemic cytokine damage. N.B. The sequence data for this patent did not		
CC	form part of the printed specification, but was obtained in electronic		
XX	format directly from WIPO at ftp.wipo.int/pub/published pct sequences		

XX	SQ	Sequence 108 AA;
 Alignment Scores:		
Pred. No.:	7,52e-30	Length: 108
Score:	494.00	Matches: 94
Percent Similarity:	89.81%	Conservative: 3
Best Local Similarity:	87.04%	Mismatches: 11
Query Match:	10.54%	Indels: 0
DB:	5	Gaps: 0
 US-09-787-016A-1 (1-2610) x ABP05496 (1-108)		
Qy	1342	TCTAGCGAAGACCAAGGGATGAAGGTTGAGAAGCTGCAAATCCAAAGTGCACAAG 1401
Dd	1	Ser-GlyGluAspHisGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLys 20
Qy	1402	AGAAACTCAAGATCTTCAGCCTGTGATAGAGGCCCTGGTGCCCTCAAATAATTATGGC 1461
Dd	21	LysLeuLysIlePheGlnProValGluAlaProGlyAlaProLysCysIleGly 40
Qy	1462	CCGGTGTGTTCACCTGCGCAGCCCCGACTCGGTGTAATGCAGTAATGACTGTATCCTC 1521
Dd	41	ProGlyCysSerSerValAlaGlnProAspSerValTyxCysSerAsnAspCysIleLeu 60
Qy	1522	AACACGCCCGCCGCGACAATGAAGTTCTTAAGCTCAGGTAAAGAACAGAGAAGCCAAAAGCCT 1581
Dd	61	LysHisAlaAlaAlaThrMetArgPheLeuSerSerGlyLysGluGlnLysThrLysPro 80
Qy	1582	AAAGAAAGATGAAGATGAAGCCAGAGAGCCAGCTCTTCCGAAATCGCGTGTCTCAGGCA 1641
Dd	81	LysGluLysValLysThrLysProGluLysPheSerLeuProLysCysSerValGlnVal 100
Qy	1642	GGTATTAAATCTCTCTCTGTCAC 1665
Dd	101	GlyIleLysIleSerSerValHis 108
 RESULT 11		
ABB71487	ID	ABB71487 standard; protein; 2016 AA.
XX	AB	ABB71487;
AC	AC	ABB71487;
XX	XX	26-MAR-2002 (first entry)
DT	DT	Drosophila melanogaster polypeptide SEQ ID NO 41253.
XX	XX	Drosophila melanogaster.
KW	KW	WO200171042-A2.
KX	KX	27-SEP-2001.
OS	OS	23-MAR-2001; 2001WO-US009231.
PN	PN	23-MAR-2000; 2000US-0191637P.
PD	PD	11-JUL-2000; 2000US-00614150.
PF	PF	(PEKE ) PE CORP NY.
PR	PR	Venter JC, Adams M, Li PWD, Myers EW;
PP	PP	WPI; 2001-655860/75.
XX	XX	N-PSDB; ABL15590.
DR	DR	New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signaling and cell-cell interactions.
PT	PT	Disclosure; SEQ ID NO 41253; 21pp + Sequence Listing; English.
PS	PS	



XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from *Drosophila*. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA  
 CC sequences (ABU1840-ABU16175) and the encoded proteins (ABB57737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences)  
 XX Sequence 2016 AA;

Alignment Scores:  
 Pred. No.: 9,27e-17 Length: 2016  
 Score: 332.00 Matches: 159  
 Percent Similarity: 32.52% Conservative: 68  
 Best Local Similarity: 22.78% Mismatches: 221  
 Query Match: 7.08% Indels: 250  
 DB: 4 Gaps: 27

US-09-787-016A-1 (1-2610) x ABB71487 (1-2016)

QY 256 GGAGCAGGTATGACGACAAAGCGACCCGAGCAATGAGGAGCACCTAAGGCCATCAAA 315  
 DB 676 GlyValGlyAsnAlaAlaAlaAspProAsn-----IleMet 689  
 QY 316 CCCACGACAAAGATTACGAAACATGGGTTTTCGAAAGACCACTATCGCCAGCGA 375  
 DB 690 AspThrAlaAsnGluAspGluIleThrAlaAspPheLeuGlnHisValValGlyLeuIle 709  
 QY 376 GAGGGCGCGGAGCGGAGGTGACCCACTGGAGCGCGCCACCCACAGCAGCAGCTG 435  
 DB 710 GluGluAspLysGlnPheGluAla----- 717  
 QY 436 GGCCTGTCTCGCGCGCAGTGGGAGGAGCCCAAGCGCACTGAGCGCTGGAGCAGTTC 495  
 DB 718 -----GluValValLysGlnVal 723  
 QY 496 CTGACACMTCCGGCGCGCGCGGAGGAGGAGCATGCTGTCTCTCCCTGGAGGATCTGGT 555  
 DB 724 LeuAla---SerThrGluProGlyThrLeuAspAlaIleValSerMetProThrSerIle 742  
 QY 556 GAGCCACGCTCTGCGCGCGCACAGACCGCGAGACGCTCCGAGGCGCAGCTG---GAA 612  
 DB 743 GluProValAspValProGlnAlaHisThrAsnLeuLeuProAsnAlaSerLeuThrGlu 762  
 QY 613 AGCGCTTCTGAGACGAGAGCGCGCCCACTGCTGCTTCCACA-----GCT 657  
 DB 763 ProAlaGlnSerMetThrSerLeuProIleAlaCysSerThrProSerArgSerValAla 782  
 QY 658 GTGAGGACGACGACGCTCTCTGAAAGGTGAAGAGGGGATGACCGATGACACACC 717  
 DB 783 AlaSerThrProProThrSerAlaLysValValArgly----- 795  
 QY 718 TCCGATAGTGACAGCGATGCGCTGACCTTCGAAAGAGCTTCAGAAATCGCTCCGAGGAG 777  
 DB 796 -----TyrGlyArgVal 799  
 QY 778 CGGGAACAGAGAGCCACTGAGAGCCCTGAAAGGATCCAGATCGCTCCGGAAGAAG 837  
 DB 800 IleTyrLeuProProIleGluAlaPro-----ThrThrArgAlaLys 813  
 QY 838 CGCGCGGAGGAG-----GGTCCCGCGAGACTGTGGGCTCCGAGGCC--- 879  
 DB 814 ArgArgAlaGlnPheProSerAlaProGlyMetAlaAlaThrSerSerSerAspAlaGly 833  
 QY 879 ----- 879  
 DB 834 AsnLeuSerPheGlyGluSerSerLeuAspAlaSerIleAsnGlnProLeuAsnThrSer 853  
 QY 880 -----AGTGACACTGTGGAGGGCGCTCTGCGCCAGTAAGCAGGAGCGCCGAGACGAT 930

DB 854 SerLeuSerAsnAspSerGlnProGlySerGlyProLysArgProAsnPro---ArgGlu 872  
 QY 931 CAGGGGGTGTGTCCAGCGCTCGGAAAGATGACAGAGAGTAGTAAAGTTGCGAGGAAGCGG 990  
 DB 873 ProSerMetAlaArgArgSerThrAlaProArgArgSerLysLysLeuAspAlaSerGln 892  
 QY 991 GTCAGGAC-----ATCAAGATGAGGAGCTGGAGACTTGGCCGCGCAGAGCT 1041  
 DB 893 AsnAsnAspProAspAlaSerGluSerGlnGluAspAsp----- 906  
 QY 1042 GAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTCGCCGAGCTCCACACAAAC 1101  
 DB 907 -----AspProAsnLysLeuTyrCysIleCysArgGlnProHisAsnAsn 921  
 QY 1102 AGGTTATGATTGCTGTGACCGCTGTGAGAAATGGTTTCATCTGGCGATTCTGGGCATT 1161  
 DB 922 ArgPheMetIleCysCysAspLeuCysGluAspTyrPheHisGlyThrCysValGlyVal 941  
 QY 1162 TCTGAGGCTCGAGGAGGCTTTTGAAGAGGAATGGGAGAGACTATATCTGCCCAACTGC 1221  
 DB 942 ThrLysAlaMetGlyThrAspMetGluAsnLysGlyIleAspTyrLysCysProLysCys 961  
 QY 1222 ACCATTCTCAAGTCGAGATGAGACTCATTCAAGAACCGCAGATCAGCAGAGAGTAA 1281  
 DB 962 -----ValLysArgGlnGluArg 968  
 QY 1282 TCGAGACCTGGAGATGCTGATGCGACCGATTGTACAAATATAGAACAAATAGACGAGAAG 1341  
 DB 969 SerGlnProArgIleThrAspMetLeuValThrArgProThrThrGlnProGluGlnArg 988  
 QY 1342 TCTAGCGAAGACCAAGGGATA-----AAGGGTAGAATTGAGAAAGCTCGAAATCCAAGT 1395  
 DB 989 ProSerGluThrLysValLeuThrThrThrAlaGluIleValGlnValAlaAlaProSer 1008  
 QY 1396 GGC---AAGAAAGAACTCAAGATCTCCAGCGCTGTGATAGAGCGCT----- 1440  
 DB 1009 AlaProArgArgThrLeuProValValLeuThrValAlaSerSerProMetArgIlePro 1028  
 QY 1441 -----GGTGCCTCAAAA----- 1452  
 DB 1029 MetAlaLysProAlaLysLysPheProThrGlyAlaIleSerHisGlnGlnGlnGln 1048  
 QY 1453 -----TGATTGCGCGCGCGG----- 1467  
 DB 1049 LeuAsnPheIleArgLeuGlyProSerProGlyLysArgIleSerGluThrLeuCysVal 1068  
 QY 1468 ---TGCTGTACGTGCGCGCAGCGCGACTCGGTGTACTGCGAGTAATGACTGTATCTCAAA 1524  
 DB 1069 ValCysLysArgProAlaSerThrSerSerValTyrCysGlyGluCysIleArgLys 1088  
 QY 1525 CACGCC-----GCA 1533  
 DB 1089 TyrAlaGlnSerAlaIleGlnAlaHisAlaAlaThrLysGlyProLeuProGlnAsnAla 1108  
 QY 1534 GGCACATAGCTTCTAGCTCAGGTAAAGACAGAGCCAAAGCCTAAAGAAAAGATG 1593  
 DB 1109 GlyAlaGlnSerLeuLeuAsnSerPheAspAlaLysAsnLysLysLysAspLeu 1128  
 QY 1593 ----- 1593  
 DB 1129 PheGluAspValLeuArgGlnAlaAspThrValSerLysValGluArgIleAsnValPhe 1148  
 QY 1593 ----- 1593  
 DB 1149 GluArgLysSerGlyArgValIleThrGlyHisMetAlaProSerSerAlaHisGlnPheArg 1168  
 QY 1594 AAGATCAAGCCAGAGAGAGCCAGT-----CTTCCGAAATGCGGTGCTCAGGAGGCT 1644  
 DB 1169 LysTyrLeuGlnGluAsnProSerPheGluValLeuProSerGlyThrValGlnSer--- 1187  
 QY 1645 ATTAATAATCTTCTGTGTCACAGAGA-----CCAGCTCCAGAAAAAAGAGACC 1695  
 DB -----

1188 -----AlaAspAlaGluLysArgLeuLeuLysGlyAlaProGluAlaAlaThrSer 1204  
1696 ACAGTGAAGAAGCAGTGGTCCCTCGCGGAGT-----GAAGCACTCGGAGGAA 1749  
1205 ThrSerGluProAlaValLeuGlyValAlaLysGlyProGluGlyProAlaLysLeu 1224  
1750 GCAGCTTGTGAGCAGCAGCGCGTCTGCGGCGGCGGATCACAATTACATGAGTAAG 1809  
1225 SerHisProGlnAsnThrThrValGlnAlaSerHisGlnLeuGlyIleSerSerValArg 1244  
1810 CCA-----GAAAGACAGTCT-----CCTCCCTCGCGGTCA 1839  
1245 ProLeuAlaLysLysAspLysGluLysThrProThrValGlnAlaProThrProAsn 1264  
1840 CTGTTGTATAATGTATGATATCATCCTAGGCGTTGGCTCTCGGACCCCTCCCGT 1893  
1265 -----ArgIleAlaGlyLysProGluProValArg 1275

RESULT 12  
AAW78908  
ID AAW78908 standard; protein; 1539 AA.  
XX  
AC AAW78908;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human protein SEQ ID NO 1570.  
XX  
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
XX  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
XX  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
XX  
KW nervous system disorder; arthritis; inflammation.  
XX  
OS Homo sapiens.  
XX  
PN WC200157190-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 05-FEB-2001; 2001W0-US004098.  
XX  
PR 03-FEB-2000; 2000US-00496914.  
XX  
PR 27-APR-2000; 2000US-00560875.  
XX  
PR 20-JUN-2000; 2000US-00598075.  
XX  
PR 19-JUL-2000; 2000US-00620325.  
XX  
PR 01-SEP-2000; 2000US-00654936.  
XX  
PR 15-SEP-2000; 2000US-00663561.  
XX  
PR 20-OCT-2000; 2000US-00693325.  
XX  
PR 30-NOV-2000; 2000US-00728422.  
XX  
PA (HVSE-) HVSEQ INC.  
XX  
PI Tang YT, Liu C, Drmanac RT, Saundt V, Zhou P, Xu C, Cao Y;  
XX  
PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren P, Chen R, Wang ZW;  
XX  
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
XX  
XX WPI; 2001-476283/51.  
XX  
XX N-PSD3; AAK52041.  
XX  
XX Nucleic acids encoding polypeptides with cytokine-like activities, useful  
XX  
XX in diagnosis and gene therapy.  
XX  
XX Claim 20; Page 3896-3899; 6221pp; English.  
XX  
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
XX  
XX encoded polypeptides (AAW78323-AAW80302) that exhibit activity relating to  
XX  
XX cytokine cell proliferation or cell differentiation or which may induce  
XX  
XX production of other cytokines in other cell populations. The  
XX  
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or  
XX  
XX peptide therapy. The polypeptides have various cytokine-like activities,  
XX  
XX e.g. stem cell growth factor activity, haematopoiesis regulating  
XX  
XX activity, tissue growth factor activity, immunomodulatory activity and  
XX  
XX activin/inhibin activity and may be useful in the diagnosis and/or

CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111  
CC (AAK52582) and 3666 (AAW80020) are omitted as the relevant pages from the  
CC sequence listing were missing at the time of publication  
XX  
SQ Sequence 1539 AA;  
Alignment Scores:  
Pred. No.: 5,35e-09 Length: 1539  
Score: 232.00 Matches: 94  
Percent Similarity: 36.43% Conservative: 63  
Best Local Similarity: 21.81% Mismatches: 154  
Query Match: 4.95% Indels: 120  
DB: 17 Gaps: 4  
US-09-787-016A-1 (1-2610) x AAW78908 (1-1539)  
QY 538 TCCTCTGGAGGAT---TCGTGTGAGCCACGCTCTCCCGCCACAGAGCCGAGACAGCC 594  
DB 95 SerLeuSerAspLysSerHisAlaHisProGlyCysLeuLysGluProHisProAla 114  
QY 595 TCCGAGGCGAGCGTGGAAAGCGCTTCTGAGACCAGA---AGCGGCCCGCAGTCTGCTCC 651  
DB 115 GlnThrGlyHisValSerHisSerGlnLysGlnCysHisLysProGlnGlnAla 134  
QY 652 ACAGCTGTGAAGGAACGACGACGCTCTTCTGAAAAGTGAAA---GGAGGGGATGACCCAC 708  
DB 135 ProAlaMetLysThrAsnSerHisValLysGluGluLeuHisProGlyValGluHis 154  
QY 709 GATGACATCTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 756  
DB 155 -----PheLysGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 170  
QY 757 CAGAAATCGCTTCGCGAGGAAGCGGAAACAGGAGGCCACTGAGAGCGCCCTGAAAGGGATC 816  
DB 171 GlnProArgGlnArgArgSerSerLysSerPheSerLeuAspGluProProLeuPheIle 190  
QY 817 CAGAGTCGCTCGCGGAGGAAGCGCGCGGAGGAGGCTCCCGCCGAGACTGTGGCTCCGAG 876  
DB 191 ProAspAsnIleAlaThrIleArgArgGlu-----GlySer--- 202  
QY 877 GCCAGTGACACTGTGGAGGGGCTCTCGCCAGTAAAGCAGAGCCGCGAGAACGATCATGGGG 936  
DB 202 ----- 202  
QY 937 GTTGTGTCCAGCTGGGAAAGATCAGAGAGAGATTAAGTTGGAGGGAAGCGGCTCAG 996  
DB 203 -----AspHisSerSerSerPheGluSerLysTyr----- 212  
QY 997 GACATCAAGATGAGAGCGCTGAGACTTGGCCCGCCAGCCGCTGAATGTGAGGGTTAC 1056  
DB 213 -----MetThrProSerLysGlnCys---GlyPhe 222  
QY 1057 GACCCCAAGCCCTGTATTGCAATTTGCGCCGAGCCCTCACAACAGAGTTTATGATTTC 1116  
DB 223 -----CysLysLysProHisGlyAsnArgPheMetValGly 234  
QY 1117 TGTGACCGCTGTGAAGATGGTTTATGCGGATTTGTTGGGCAATTTCTGAGCTCCGAGG 1176  
DB 235 CysGlyArgCysAspAspTrpPheHisGlyAspCysValGlyLeuSerLeuSerGlnAla 254  
QY 1177 AGGCTTTTGGAAAGGAATGGGAAGACTATCTGCCAAACTGACCACTTCTGCAAGTG 1236  
DB 255 GlnGlnMetGlyGluGluAspLysGluTyrValCysValLysCysCys-----Ala 271  
QY 1237 CAGGATGAGACTCATTTAGAAACGGCAGAT-----CAGCAGGAGGCT 1278  
DB 272 GluGluAspLysLysThrGluIleLeuAspProAspThrLeuGluAsnGlnAlaThrVal 291  
QY 1279 AATGAGACCTGGAGATCTGATGCGACCGGATTGTACAAGTATAGGAACAATAGAGCAG 1338  
DB 292 GluPheHisSerGlyAspLysThr---MetGluCysGluLysLeuGlyLeuSerLysHis 310



[illegible]

**XX**

03-FEB-2000; 2000US-00496914.

**PR**

27-APR-2000; 2000US-00508075.

**PR**

20-JUN-2000; 2000US-00598075.

**PR**

19-JUL-2000; 2000US-00620325.

**PR**

01-SEP-2000; 2000US-00654936.

**PR**

15-SEP-2000; 2000US-00663561.

**PR**

20-OCT-2000; 2000US-00693325.

**PR**

30-NOV-2000; 2000US-00728422.

**XX**

(HYSE-) HYSEQ INC.

**PA**

Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;

**PI**

Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;

**P1**

Xue AJ, Yang Y, Wejрман T, Goodrich R,

**XX**

WI; 2001-476283/51.

**DR**

N-PSDB; AAK53025.

**XX**

Nucleic acids encoding polypeptides with cytokine-like activities, useful  
in diagnosis and gene therapy.

**PT**

**XX**

Claim 20; Page 376-377; 5221pp; English.

**PS**

The invention relates to polymynucleotides (AAK51456-AAK53435) and the  
CC encoded polypeptides (AAM78323-AAM90302) that exhibit activity relating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polymynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC actinin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52591), 2111  
CC (AAK52582) and 3666 (AAM90020) are omitted as the relevant pages from the  
CC sequence listing were missing at the time of publication

**XX**

SQ Sequence 1766 AA;

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	Gaps:	DG:
5.62e-09	232.00	36.43%	21.81%	4.95%	4	
Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:	
1766	94	63	154	120	17	

US-09-787-016A-1 (1-2610) x AAW79892 {1-1766}

**OY**

538 TCCTGTGGAGAT---TCTGGTGAGCCCACGCCTCGTCGCCCCCCCACCAGACGCCGAGACGCC 594

**Db**

322 SerLeuSerAspLysSerHisAlaHisProGlyCysLeuLysGluProHisHisProAla 341

**OY**

595 TCCGAGGCCAGCGCTGAAGAAGCGTTCTGAGAACAGACA---AGCGGCCGCCCACTGTCTGTC 651

**Db**

342 GluThrGlyHisValSerHisSerSerGlnLysGlnCysHisLysProGlnGlnGlnAla 361

**OY**

652 ACAGCTGTGAAGAAACGACCGACTCTTCTGAATAAGGTGA-- --GGAGGGGATGACCATC 708

**Db**

362 ProAlaMetIleThyrAnSsrHisValLysSGlUglUeuGlHIsProGlyValGIUHIS 381

**OY**

709 GATGACACCTCCGATAGTAGCACGATGGCTGCACCTTAAGAG-----CTT 756

**Db**

382 -----PheLysGluGluAbSPlysLeuLysLeuLysLysProGluLysAsnLeu 397

**OY**

757 CAGAATCCCTTCGAGGAACGGGACACGAGGCCCATCGAGAGGCCCTTGAAGGGATC 816

**Db**

398 GlnProArgGlnArgSerSerLysSerPheSerLeuASPGLUProProLeuPheIle 417

**OY**

817 CAGAGTCGCTCGGAGAGAGCGCCGGAGAGGGTCCCGCCGAGACTGTGGCTCCGAG 876

**Db**

418 ProASpnileAleThrIlea-garggllu-----Glycer--- 429

```
QY 877 GCCAGTGCACACTGTGGAGGCGTCTCTGCCCAAGCAGGAGCCGAGACGATCAGGGG 936
Db 429 -----
QY 937 GTTGTGTCCTCCAGCTGGGAAGATGACAGAGAGTAAGTTGAGGAGGAAGCGGCTCAG 996
Db 430 -----
QY 997 GACATCAAGATGAGGAGCCTGGAGACTTGGGCGGACCGAAGCCTGAATGTGAGGGTTAC 1056
Db 440 -----
QY 1057 GACCCCAACCCCTGTATTCCTCCTGCGCCAGCCTCAACAACAGCTTTATGATTTGC 1116
Db 450 -----
QY 1117 TGTGACCGCTGTGAAGATGTTTGTGTGGCGATTGTGGGCTTTCTGAGGCTCGAGGG 1176
Db 462 CysGlyArgCysAspAspThrPheHisGlyAspCysValGlyLeuSerLeuSerGlnAla 481
QY 1177 AGCTTTTGAAGGAATGGGAGACTATATCTGCCAACTGCACCATTTCTGCAAGTG 1236
Db 482 GlnGlnMetGlyGluGluAspLysGluTyValCysValLysCysCys-----Ala 498
QY 1237 CAGATGAGACTCATTTCAGAAACGGCAGAT-----CAGCAGGAGAGCT 1278
Db 499 GluGluAspLysLysThrGluLeuLeuAspProAspThrLeuGluAsnGlnAlaThrVal 518
QY 1279 AATGTGAGACTGAGATGCTGTATGGCACCGATGTACAAAGTATAGAAACAATAGACGAG 1338
Db 519 GluPheHisSerGlyAspLysThr--MetGluCysGluLysLeuGlyLeuSerLysHis 537
QY 1339 AAGTCTAGCGAA-----GACCAAGGATAAAGGTAGAGTGAATGAGAAAGCT 1383
Db 538 ThrThrAsnAspArgThrLysTyrlleAspAspThrValLysHisValLysIleLeu 557
QY 1384 GCAATCCAAAGTGCAGAGAAACTCAAGATCTTCAGAGCTGTGTAGAGGCGCCTGGT 1443
Db 558 LysArgGluSerGlyGluGlyArg----- 565
QY 1444 GCCTCAAAATGTATGTGCGCCCGGGTGTGTACGTGGCGGAGCCGAGCTCGGTGTACTGC 1503
Db 566 -----
QY 1504 AGTAATGACTGTATCTCAAAACGCGCGCAGCAGCAATGAATGCTTCAAGCTCAGGTAAA 1563
Db 571 ArgAspAsnGluIleLysLysThrGlnLeuAlaProLeuArgLysMetGlyGlnProVal 590
QY 1564 GACAGAGAACCAAGCCTAAAGAAAGATGAAGATGAAGCAGAGAACCCAGCTTCCG 1623
Db 591 LeuProArgArgSerSerGluGlySerGluLysIleProLysGluSerThrThrVal 610
QY 1624 AATGCGGTCTCAGGCGAGTATTAATCTCTCTGTGCAAGAGACAGCTCCAGAA 1683
Db 611 ThrCys-----ThrGlyGluLysAlaSerLysProGlyThrHisGluLysGlnGlu 627
QY 1684 AAAAAGAGACACACAGTAAGAGAGGAGTGTGTGCTCCTCGCGGAGTGAAGACTCGGG 1743
Db 628 MetLysLysLysLysValGluLysGlyVal-----LeuAsn 639
QY 1744 AAGAGAGCAGCTGTGAGACAGCAGCAGCCTGTG 1776
Db 640 ValHisProAlaAlaSerAlaSerLysProSer 650
RESULT 15
AAM41020
ID AAM41020 standard; protein; 1766 AA.
XX
AC AAM41020;
XX
DT 22-OCT-2001 (first entry)
XX
```

## Human polypeptide SEQ ID NO 5951.

```
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia.
XX Homo sapiens. OS
XX WO200153312-A1.
XX 26-JUL-2001.
XX 26-DEC-2000; 2000WO-US034263.
XX 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX WPI: 2001-442253/47.
DR N-PSDE; AAI60176.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders such
PT as central nervous system injuries.
XX
XX Example 2; SEQ ID NO 5951; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC encoded polypeptides (AAI38642-AAI42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, anyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
XX part of the printed specification
```

## Sequence 1766 AA;

Alignment Scores:			
Pred. No.:	5,62e-09	Length:	1766
Score:	232.90	Matches:	94
Percent Similarity:	36.43%	Conservative:	63
Best Local Similarity:	21.81%	Mismatches:	154
Query Match:	4.95%	Indels:	120
DB:	4	Gaps:	17

US-09-787-016A-1 (1-2610) x AAM41020 (1-1766)

QY	538	TCCTGGAGGAT---TCTGTGAGCCCGCTCTCCCGCCGAGACGCGGAGACGCC	594
DB	322	SerLeuSerAspLysSerHisAlaHisProGlyCysLeuLysGluProHisProAla	341

Qy	595	TCCGAGGGCAGCGTGGAAAGCGCTTCTGTGACACCAGA---AGCGCCGCCCAAGTCTGCTTCC	651
Db	342	GlnThrGlyHisValSerHisSerGlnLysSlnCysHisLysProGlnGlnAla	361
Qy	652	ACAGCTGTGAAGAAACGACCGAGCTCTTCTGAAAGGTGAAA---GGACGGGATGACCC	708
Db	362	ProAlaMetLysThrAenSerHisValLysGluGluLeuGluHisProGlyValGluHis	381
Qy	709	GATGACACCTCCGATAGTCACAGCATGGCGCTGACCTTGAAGACG---CCT	756
Db	382	-----PheLysGluGluAspLysLeuLysLysProGluLysAsnLeu	397
Qy	757	CAGAAATCGCTTCCAGAAAGCGGGAACAGGACCCACTCAGAGGCCCTTGAAGGGATC	816
Db	398	GlnProArgGlnArgArgSerSerLysSerPheSerLeuAspGluProLeuPheile	417
Qy	817	CAGAGTCGCTTCGCGAAAGACGCGCGGAGGAGGGTCCCGCGAGACTGTGGCTCCGAG	876
Db	418	ProAspAsnIleAlaThrIleArgArgGlu-----GlySer---	429
Qy	877	GCCAGTCGCACTGTGGAGGGCGTCTCCCGCAGTAAGCAGGAGCCCGAGAAGCATCAGGG	936
Db	429	-----	429
Qy	937	GTTGTGTCGCCAGCTCGGAAGATCACAGAGAGTAAGTTGGAGGAAAAGCGGCTCAG	996
Db	430	-----AspHisSerSerSerPheGluSerLysTyr-----	439
Qy	997	GACATCAAGAATGAGGAGCGCTGGAGACTTGGCGCGACCGAAGCTGAATGTGAGGGTTAC	1056
Db	440	-----MetTrpThrProSerLysGlnCys---GlyPhe	449
Qy	1057	GACCCCAAGCCCTGTATTGCAATTGCGCGCCAGCTCACACACAGGTTTATGATTTGC	1116
Db	450	-----CysLysLysProHisGlyAsnArgPheMetValGly	461
Qy	1117	TGTACCCCTGTGAAGAATGTTTCATCGGCAATTGTGTGGGCATTTCTGAGGCTCGAGGG	1176
Db	462	CysGlyArgCysAspAspTrpPheHisGlyAspCysValGlyLeuSerLeuSerGlnAla	481
Qy	1177	AGGCTTTTGGAAAGAAATGGGGAAGACTATATCTGCCAAACTGCACCATTCTGCAAGTG	1236
Db	482	GlnGlnMetGlyGluGluAspLysGluTrpValCysValLysCysCys-----Ala	498
Qy	1237	CAGGATCAGACTCATTCAGAAACGGCAGAT-----CAGCAGGAAAGCT	1278
Db	499	GluGluAspLysLysThrGluIleLeuAspProAspThrLeuGluAsnGlnAlaThrVal	518
Qy	1279	AAATGCGAGACCTCGAGATGCTGTGCGACCGCATTTGTACAGATATAGAAACATAGACGAG	1338
Db	519	GluPheHisSerGlyAspLysThr---MetGluCysGluLysLeuGlyLeuSerLysHis	537
Qy	1339	AAGTCTAGCGAA-----GACCAAGGGATAAAGGGTAGAATTTGACAAAGCT	1383
Db	538	ThrThrAsnAspArgThrLysTyrlleAspAspThrValLysHisLysValLysIleLeu	557
Qy	1384	GCAATCCAACTGGCAAGAAAGAACTCAAGATCTTCCAGCTGTGTGATAGAGCGCCCTGGT	1443
Db	558	LysArgGluSerGlyGluGlyArg-----	565
Qy	1444	GCCTCAAAATGATTGGCCCGCGGTGCTGTCACGTGCGCAGCCGACTCGGTGCTACTGC	1503
Db	566	-----AsnSerSerAspCys	570
Qy	1504	AGTATGACTCTATCTCTCAAAACGCGCGCAGCGCAATAGTATTCTCAAGCTCAGGTAAA	1563
Db	571	ArgAspAsnGluIleLysLysTrpGlnLeuAlaProLeuArgLysMetGlyGlnProVal	590
Qy	1564	GAAACGAGCCAAAGCCTAAAGAAAGATGAAGATGAGCCAGAGAGAGCCAGCTTCCG	1623
Db	591	LeuProArgArgSerSerGluGluLysSerGluLysIleProLysGluSerThrThrVal	610
Qy	1624	AAATGGCGGTGCTCAGCGAGGTATTAAATCTCTCTGTGCAACAGAGACAGCTCCAGAA	1683

Db 611 ThrCys-----ThrGlyGluLysAlaSerLysProGlyThrHisGluLysGlnGlu 627  
 QY 1684 AAAAAGAGACACACAGTGAAGAAGCGCAGTGGTGGTCCCTCGCGGAGTGAAGCACTCGGG 1743  
 Db 628 MetLysLysLysLysValGluLysGlyVal-----LeuAsn 639  
 QY 1744 AAGGAACGACGTTGTGAGACGACGACGCCCTCG 1776  
 Db 640 ValHisProAlaAlaSerAlaSerLysProSer 650

Search completed: April 28, 2004, 10:54:06  
Job time : 153 secs